BIOMETRIKA

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

IN CONSULTATION WITH

VOLUME XXXIV

1947

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

First printed in Great Britain at the University Press, Cambridge

Reprinted by offset-litho by Percy Lund Humphries & Co. Ltd

CONTENTS OF VOLUME XXXIV

Memoirs

Miscellanea

Vol. XXXIV. Parts I and II January, 1947

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON

AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

Reprinted by offset-litho 1953

[*Issued* 11 February 1947]

THE VARIANCE OF THE OVERLAP OF GEOMETRICAL FIGURES WITH REFERENCE TO A BOMBING PROBLEM

BY F. GARWOOD, PH.D.

1. INTRODUCTION

The present paper deals with a particular problem arising in the mathematical study of bombing. Briefly, the general problem is that of predicting the over-all effects of a bombing attack carried out under given conditions against a given target, and the mathematical treatment involves various simplifying assumptions concerning these conditions.

In the type of problem considered here, attention is centred on the total plan area of damage caused *to a single building by bombs falling independently and at random over a* larger area containing the building. It is assumed that each bomb damages all that part of the building contained within a circle of fixed size centred at the bomb (a square damage area is also considered), while the building has a simple plan outline, such as a rectangle or a circle. The area of damage of two or more adjacent bombs is merely the area covered by the circles. The theoretical problems dealt with are those of estimating the variance of the amount of damaged area (the estimation of the mean or expected damage presents little difficulty). It would be more satisfactory to obtain the complete frequency distributions, but this has so far not been achieved, nor has it been possible to obtain explicit formulae for the 3rd and 4th moments.

As there may be applications of the problems to fields completely different from those of bombing studies, and as they are problems which involve essentially the concepts of geometry and of probability, it is convenient to express them entirely in these terms.

We thus have problems of the following type. A number of circles are placed at random on a plane so that each one has some or all of its area inside a fixed square. What are the mean and variance of the area of the square covered by the circles? The fundamentals of this type of problem have been studied by Robbins (1944), and Bronowski & Neyman have dealt with another particular case.* Robbins's results enable us to deal with geometrical figures other than circles and squares, and also to deal with cases where the number, position and orientation of the 'covering' figures follow probability laws other than the simple ones implied in the above example.

2. ROBBINS'S THEOREM

In leading up to his theorem, Robbins uses the concept of a random measurable subset X of n-dimensional Euclidean space E_n. He defines the function $g(x, X)$ for every point x of E_n and for every X as equal to 1 for $x \in X$ and zero elsewhere. This theorem is then as follows:

Let X be a random Lebesgue measurable subset of E_n, with measure $\mu(X)$. For any point x of E_n let $p(x) = Pr(x \in X)$. Then, assuming that the function $g(x, X)$ is a measurable function of the pair (x, X), the expected value of the measure of X will be given by the Lebesgue integral of the function $p(x)$ over E_n.

* *Note by Editor.* This paper was received for publication in September 1945; Dr Garwood has asked me to add the following note in proof. "The author had the privilege of seeing the work of Bronowski & Neyman in proof. This paper was then submitted, after which their work was published (1945) together with a second article by Robbins (1945), who has solved, among others, some of the problems dealt with in this paper, as acknowledged in later footnotes."

Robbins generalizes this result to obtain the mth moment of the measure of X; this is the integral of the function $p(x_1, x_2, \ldots, x_m)$ over E_{mn}, where

$$p(x_1, x_2, \ldots, x_m) = Pr(x_1 \in X \text{ and } x_2 \in X \ldots \text{ and } x_m \in X). \tag{1}$$

It is useful to give a simple non-rigorous proof of this result. Suppose the space E_n to be divided into an enumerably infinite set of small elements $\omega_1, \omega_2, \ldots$. If we assume that any particular subset X can be made up of a selection of the ω's, then

$$\mu(X) = \lambda_1\omega_1 + \lambda_2\omega_2 + \ldots,$$

where ω_i is here used also as the measure of the element ω_i, and where the λ's, appropriate to this particular X, are 0 or 1. Hence

$$\{\mu(X)\}^m = \sum_p \sum_q \ldots \lambda_p \lambda_q \ldots \omega_p \omega_q \ldots,$$

where each summation is over the whole of E_n, and there are m such summations. Thus

$$\exp\{\mu(X)\}^m = \sum_p \sum_q \ldots \omega_p \omega_q \ldots \exp(\lambda_p \lambda_q \ldots).$$

But the expectation of $\lambda_p \lambda_q \ldots$ is the probability that the elements $\omega_p, \omega_q, \ldots$ are in X, and on proceeding to the limit the desired result is obtained.

The verification of this result in the case, say, of the 2nd moment of a linearly distributed variate, is instructive. Thus suppose x is a variate with a probability function $F(x)$, i.e. the probability of obtaining a value $\leqslant x$ is given by the measurable function $F(x)$, where

$$F(-\infty) = 0 \quad \text{and} \quad F(\infty) = 1.$$

Define X as the interval from 0 to x; then the expectation of the square of the measure of X is the 2nd moment of x. To use Robbins's theorem we used the probability $p(x_1, x_2)$ that a given pair of values x_1 and x_2 both lie in the interval $0, x$ chosen at random. Using coordinate axes Ox_1, Ox_2, this probability is zero in the 2nd and 4th quadrants, since O, x cannot contain two points x_1 and x_2 of opposite signs. In the region A (see Fig. 1), where $x_1 > x_2 > 0$, the two values x_1 and x_2 are both in O, x if $x > x_1$, and the probability of this is $1 - F(x_1)$. Thus in A,

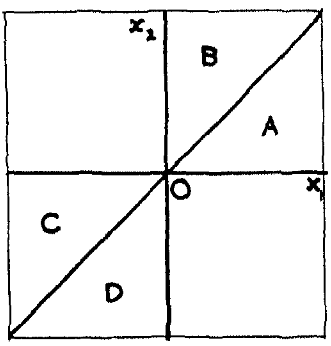


Fig. 1.

$$p(x_1, x_2) = 1 - F(x_1).$$

Similarly in B $\quad p(x_1, x_2) = 1 - F(x_2)$,

while in C $\quad p(x_1, x_2) = F(x_1)$

and in D $\quad p(x_1, x_2) = F(x_2)$.

The integral of $p(x_1, x_2)$ over A is seen to be

$$\int_0^\infty x_1[1 - F(x_1)]\,dx_1,$$

while the total integral of $p(x_1, x_2)$ over the whole plane is

$$2\int_0^\infty x[1 - F(x)]\,dx - 2\int_{-\infty}^0 xF(x)\,dx.$$

A single integration by parts then leads to

$$\int_{-\infty}^\infty x^2\,dF(x),$$

which is the 2nd moment of x, as required.

3. APPLICATION OF ROBBINS'S THEOREM TO OVERLAP PROBLEMS

We shall be concerned with cases in E_2 where the subset X is the part of a fixed area A in the plane which is covered by a number of areas C dropped independently and at random on the plane. We suppose A to be the interior of a simple closed curve, while each C is the interior of another curve. The area C has a reference point Q (conveniently called its centre) and a reference line, and it is assumed that there is a frequency distribution $\phi(x, y, \theta)$ of the position (x, y) of Q and of the inclination θ of the reference line to a fixed direction in E_2. $\phi(x, y, \theta)$ can be assumed to be zero outside an area T, i.e. the points Q are distributed inside T. (In the applications the angle θ will be constant and the areas C will be equally likely to fall anywhere over T, so that we can write $\phi(x, y, \theta) = 1/T$.) Another chance variable is k, the number of areas C; its distribution can be defined by the series $p_0, p_1, p_2, \ldots, p_k, \ldots$ (which are the probabilities of $0, 1, 2, \ldots, k, \ldots$ C's falling on T), or by the probability generating function $G(u)$, where

$$G(u) \equiv p_0 + p_1 u + p_2 u^2 + \ldots + p_k u^k + \ldots. \tag{2}$$

Finally, it is more convenient to consider the moments of the area $Y = A - X$, i.e. the area of A (we can use the symbols Y, etc. for either the sets or their areas) not covered by the C's. Evidently the variances of X and Y are equal.

To obtain the 1st moment of Y, we need first the probability $p(x_1, y_1)$ that a point (x_1, y_1) of A will belong to Y, i.e. of (x_1, y_1) not being covered by a C. Now (x_1, y_1) will not be covered by a particular C falling at an inclination θ if the centre $Q(x, y)$ falls outside an area $\bar{C}(x_1, y_1, \theta)$ obtained by centring the C at (x_1, y_1) and rotating it through 180°. If the part of T exterior to this area is called $T - \bar{C}(x_1, y_1, \theta)$, and if we allow all inclinations, the probability of this occurring is

$$q(x_1, y_1) = \int_0^{2\pi} \iint_{T - \bar{C}(x_1, y_1, \theta)} \phi(x, y, \theta)\, dx\, dy\, d\theta. \tag{3}$$

If k C's are dropped independently, the probability is given by $q^k(x_1, y_1)$, so that the total probability of (x_1, y_1) belonging to Y is

$$p(x_1, y_1) = \sum_{k=0}^{\infty} p_k q^k(x_1, y_1) = G\{q(x_1, y_1)\}. \tag{4}$$

The 1st moment of Y is thus, in the case of k C's,

$$\mu_1'(Y) = \iint_A q^k(x_1, y_1)\, dx_1\, dy_1, \tag{5}$$

and in the general case

$$\mu_1'(Y) = \iint_A G\{q(x_1, y_1)\}\, dx_1\, dy_1. \tag{6}$$

The 2nd moment is obtained by a similar process; we require the probability $p(x_1, y_1, x_2, y_2)$ that neither of two points (x_1, y_1) and (x_2, y_2) is covered by a C. Corresponding to these two points and an inclination θ the permissible region in which each centre Q can fall is

$$T - \bar{C}(x_1, y_1, \theta) - \bar{C}(x_2, y_2, \theta) \equiv T - \bar{C}_1 - \bar{C}_2.$$

Thus for one C the probability is

$$q(x_1, y_1, x_2, y_2) = \int_0^{2\pi} \iint_{T - \bar{C}_1 - \bar{C}_2} \phi(x, y, \theta)\, dx\, dy\, d\theta, \tag{7}$$

giving in the case of k C's,

$$p(x_1, y_1, x_2, y_2) = q^k(x_1, y_1, x_2, y_2) \quad \text{and} \quad p(x_1, y_1, x_2, y_2) = G\{q(x_1, y_1, x_2, y_2)\}$$

in the general case. Thus the 2nd moment

$$\mu_2'(Y) = \int_A\int\int_A\int q^k(x_1, y_1, x_2, y_2)\,dx_1dy_1dx_2dy_2 \quad \text{for } k\ C\text{'s} \tag{8}$$

and

$$\mu_2'(Y) = \int_A\int\int_A\int G\{q(x_1, y_1, x_2, y_2)\}\,dx_1dy_1dx_2dy_2 \tag{9}$$

in the general case. In general, for the mth moment, the probability that $(x_1, y_1) \dots (x_m, y_m)$ are not covered by k C's is

$$q^k(x_1, y_1, x_2, y_2, \dots, x_m, y_m),$$

where

$$q(x_1, y_1, x_2, y_2, \dots, x_m, y_m) = \int_0^{2\pi}\int\limits_{T-\bar{C}_1-\bar{C}_2-\dots-\bar{C}_m}\int \phi(x, y, \theta)\,dx\,dy\,d\theta, \tag{10}$$

and $T-\bar{C}_1-\bar{C}_2\dots-\bar{C}_m$ is the area of T outside C's centred at (x_1, y_1), (x_2, y_2), …, (x_m, y_m) and rotated through 180°.

Thus the mth moment is equal to

$$\left.\begin{aligned} \mu_m'(Y) &= \int\limits_A\int\dots\int\limits_A\int q^k(x_1, y_1, x_2, y_2, \dots, x_m, y_m)\,dx_1dy_1dx_2dy_2\dots dx_mdy_m \\ &\text{in the case of } k\ C\text{'s, or} \\ &= \int\limits_A\int\dots\int\limits_A\int G\{q(x_1, y_1, x_2, y_2, \dots, x_m, y_m)\}\,dx_1dy_1dx_2dy_2\dots dx_mdy_m \end{aligned}\right\} \tag{11}$$

in the general case.

4. Uniform distribution of covering areas at constant inclination

As mentioned above, in the cases with which we shall be dealing, the areas C are equally likely to fall anywhere over T, and the angle θ is constant. The function $\phi(x, y, \theta)$ can be put equal to $1/T$ for points of T and zero outside; the variable θ, and integration with respect to it, may be omitted.

The function $q(x_1, y_1)$ is the fraction of T not covered by a C centred at (x_1, y_1) and rotated through 180°, and in general $q(x_1, y_1, x_2, y_2, \dots, x_m, y_m)$ is the fraction of T outside m C's centred at (x_1, y_1), (x_2, y_2), …, (x_m, y_m) and rotated through 180°.

Instead of the variate Y we can consider Y/A, i.e. the fraction of A not covered by k C's, and to obtain its mth moment we divide $\mu_m'(Y)$ by A^m. Also the quantity $dx_1dy_1\dots dx_mdy_m/A^m$ is the probability of obtaining m centres in the elements of area $dx_1, dy_1, \dots, dx_mdy_m$ of A if these centres are uniformly distributed over A.

We thus obtain the following result from (11): *the mth moment of the fraction of A not covered by k C's with their centres falling at random on T is equal to the kth moment of the fraction of T not covered by m C's with their centres falling at random on A and rotated through 180°.*

In the case $k = m = 1$ we can express this in a slightly different way if we (i) deal with the area common to the two areas concerned, (ii) regard all orientations as possible and as equally likely, and (iii) deal with areas rather than fractions. We obtain, in fact, the following result: *the integral of the overlap of C and A, when the centre of C is taken over T and all orientations are permitted, is equal to the corresponding integral of the overlap of C and T, for all positions of the centre of C on A and for all orientations.*

In the practical cases with which we shall deal, the area A is always 'well inside' T, i.e. every point of A can be reached by a C centred somewhere in T. In such cases the formula

for the mean overlap is simple; we have $m = 1$ and the fraction of T not covered by one C is $(T-C)/T$, which is constant for all (x_1, y_1), so that its kth moment is $(T-C)^k/T^k$, i.e.

$$\mu_1'(Y/A) = \left(\frac{T-C}{T}\right)^k. \tag{12}$$

If the number of C's follows a probability generating function $G(u)$, the mean is given by

$$\mu_1'(Y/A) = G\left(\frac{T-C}{T}\right). \tag{13}$$

For the 2nd moment we are concerned with two C's centred at (x_1, y_1) and (x_2, y_2), and if their common area is $\Omega(x_1, y_1, x_2, y_2)$, we have

$$q(x_1, y_1, x_2, y_2) = \frac{T-2C+\Omega(x_1, y_1, x_2, y_2)}{T}. \tag{14}$$

The 2nd moment $\mu_2'(Y/A)$ is then the expectation of q^k or $G(q)$ for all pairs of points over A, and we no longer have a simple formula as in the case of the 1st moment. The overlap Ω, however, depends on the relative positions of the two C's, and therefore the number of variables in the integration is reduced from 4 to 2 or 1. This is illustrated in the following examples.

5. Circles falling on a fixed square

Assume A to be a square of unit side (i.e. $A = 1$), C a circle radius a and T a 'square with rounded corners', whose boundary is at a distance a outside the sides of A. Thus

$$T = 1+4a+\pi a^2 \tag{15}$$

and

$$C = \pi a^2. \tag{16}$$

It is seen that the fraction q of T outside the two circles centres (x_1, y_1) and (x_2, y_2) is a function only of the distance r between these points, and can therefore be written as $q(r)$. Hence if $\phi(r)$ is the frequency function of r, we obtain

$$\mu_2'(Y) = \int_0^{\sqrt{2}} q^k(r)\,\phi(r)\,dr, \tag{17}$$

or

$$\mu_2'(Y) = \int_0^{\sqrt{2}} G\{q(r)\}\,\phi(r)\,dr. \tag{18}$$

The area $\Omega(r)$ common to two circles radii a with centres distant r apart is

$$\Omega(r) = 2a^2(\theta - \sin\theta\cos\theta), \tag{19}$$

where

$$r = 2a\cos\theta \quad (r \leqslant 2a), \tag{20}$$

and

$$\Omega(r) = 0 \qquad (r \geqslant 2a), \tag{21}$$

and

$$q(r) = \frac{1+4a-\pi a^2+\Omega(r)}{1+4a+\pi a^2}, \tag{22}$$

where $\Omega(r)$ is given by (19), (20) and (21).

To obtain the frequency function $\phi(r)$ of r, we note that

$$r^2 = (x_1-x_2)^2+(y_1-y_2)^2, \tag{23}$$

where x_1, x_2, y_1 and y_2 are uniformly and independently distributed in the range 0, 1. The difference $\xi = |x_1 - x_2|$ follows the 'triangular' distribution

$$df = 2(1-\xi)\,d\xi, \tag{24}$$

from $\xi = 0$ to 1, so that the quantity

$$u = \xi^2 = (x_1-x_2)^2$$

follows the distribution

$$df = \frac{1-\sqrt{u}}{\sqrt{u}}\,du. \tag{25}$$

Similarly, $v = (y_1 - y_2)^2$ follows independently the same distribution

$$df = \frac{1-\sqrt{v}}{\sqrt{v}}\,dv. \tag{26}$$

The distribution of $$r = \sqrt{(u+v)}$$

is obtained by integrating the product

$$\frac{\sqrt{(1-u)}\sqrt{(1-v)}}{\sqrt{(uv)}} \tag{27}$$

of the frequencies of u and v over that part of the line $u+v=r^2$ within the square of unit side in which the point u, v can lie, and we obtain without much difficulty for the frequency function of r,

$$\phi(r) = 2r(\pi - 4r + r^2) \quad \text{for} \quad 0 < r < 1, \tag{28}$$

and $$\phi(r) = 2r(4\sin^{-1} 1/r + 4\sqrt{(r^2-1)} - r^2 - \pi - 2) \quad \text{for} \quad 1 < r < \sqrt{2}. \tag{29}$$

Thus the 2nd moment of the fraction of the unit square uncovered is given by the integral (17) or (18), where $q(r)$ is given by (22), (19), (20) and (21) and $\phi(r)$ is given by (28) and (29).

It does not appear possible to reduce the integral (17) simply to elementary functions, and quadrature must be used. The integrand has discontinuities in its first derivative at $r = 1$ and $r = 2a$, so that the integration must be carried out separately over the intervals with these as end-points.

6. Overlap of circles on fixed rectangle*

We replace the square A of the previous section by a rectangle A; for convenience we assume its sides to be $\sqrt{b}$ and $1/\sqrt{b}$, where $b > 1$, so that the ratio of the longer to the shorter side is b and the area is unity. The centres of the circles radii a are assumed to be equally likely to fall anywhere in a 'rectangle with rounded corners' T, whose boundary is at a distance a outside A, i.e.

$$T = 1 + \pi a^2 + 2a(\sqrt{b} + 1/\sqrt{b}). \tag{30}$$

To obtain the 2nd moment of the fraction of the area of A not covered, we calculate an integral similar to (17) or (18). The function

$$q(r) = \frac{T - 2\pi a^2 + \Omega(r)}{T}$$

is derived from $\Omega(r)$, which remains the same, but the frequency distribution $\phi(r)$, the distance between a pair of points chosen at random in the rectangle, is different.

The co-ordinates x_1 and x_2 are uniformly and independently distributed in the range O, $\sqrt{b}$ (if we take Ox parallel to the longer side). The distribution of $u = (x_1 - x_2)^2$ is thus seen from (25) to be

$$\begin{aligned} df &= \frac{1-\sqrt{(u/b)}}{\sqrt{(u/b)}}\frac{du}{b} \\ &= \frac{\sqrt{b}-\sqrt{u}}{b\sqrt{u}}\,du, \end{aligned} \tag{31}$$

while the distribution of $v = (y_1 - y_2)^2$ is

$$df = \frac{1-\sqrt{(bv)}}{\sqrt{(bv)}}\,b\,dv. \tag{32}$$

* This problem was solved by Robbins (1945); see footnote on p. 1.

The distribution of $r = \sqrt{(u+v)}$ is obtained by integrating the product

$$\frac{(\sqrt{b}-\sqrt{u})\,(1-\sqrt{(bv)})}{\sqrt{(buv)}}$$

over that part of the line $u+v = r^2$ within the rectangle $0 \leqslant u \leqslant b$, $0 \leqslant v \leqslant 1/b$. This gives

$$\phi(r) = \phi_1(r) = 2r[\pi - 2r(\sqrt{b}+\sqrt{(1/b)})+r^2] \quad \text{for} \quad r < 1/\sqrt{b}, \tag{33}$$

$$\left.\begin{aligned} &\phi(r) = \phi_2(r) = 2r[2\alpha - 1/b - 2r\sqrt{b}(1-\cos\alpha)] \quad \text{for} \quad 1/\sqrt{b} < r < \sqrt{b},\\ &\text{where} \qquad \alpha = \sin^{-1} 1/r\sqrt{b}, \end{aligned}\right\} \tag{34}$$

$$\left.\begin{aligned} &\text{and} \quad \phi(r) = \phi_3(r) = 2r[2(\alpha-\beta) - b - 1/b + 2r\sin\beta/\sqrt{b} + 2r\sqrt{b}\cos\alpha - r^2]\\ &\text{for} \qquad \sqrt{b} < r < \sqrt{(b+1/b)},\\ &\text{where} \qquad \beta = \cos^{-1}\sqrt{b}/r. \end{aligned}\right\} \tag{35}$$

Thus the 2nd moment of Y can be found from (19)–(21) together with (30) and (33)–(35).

7. Overlap of rectangles on a fixed rectangle

Assume that the fixed rectangle A has sides a and b and the covering rectangles C have sides α and β. The latter are assumed to be dropped with sides α parallel to the side a and with their centres anywhere inside the rectangle T, which is concentric with ab and has sides $a+\alpha$ and $b+\beta$. To calculate the 2nd moment of the fraction Y/A of A not covered by k C's, we use (18) and calculate the expectation of $q(x_1, y_1, x_2, y_2)$, the fraction of T not covered by two C's with their centres (x_1, y_1) and (x_2, y_2) falling at random in A. The area common to two C's is readily seen to depend only on the difference ξ of the x co-ordinates of their centres and on the similar difference η of their y co-ordinates. In fact, the area can be written as

$$\Omega(x_1, y_1, x_2, y_2) = [\alpha - \xi][\beta - \eta], \tag{36}$$

where the symbol* $[x]$ stands for x when $x > 0$ and is zero when $x < 0$, and we obtain

$$q(x_1, y_1, x_2, y_2) = 1 + \frac{[\alpha-\xi][\beta-\eta] - 2\alpha\beta}{(a+\alpha)(b+\beta)}. \tag{37}$$

To obtain the expectation of q^k, we need the frequency distribution of ξ and η. As in § 5, ξ is readily seen to follow the frequency distribution

$$df = \frac{2(a-\xi)}{a^2}\,d\xi \tag{38}$$

between O and a, with a similar distributon for η, and we obtain the result

$$\mu_2'(Y/A) = \frac{4}{a^2b^2}\int_0^a\int_0^b \left\{1 + \frac{[\alpha-\xi][\beta-\eta]-2\alpha\beta}{(a+\alpha)(b+\beta)}\right\}^k (a-\xi)(b-\eta)\,d\xi\,d\eta. \tag{39}$$

If the kth power be expanded, the resulting integrals are, with the exception of the first, the product of integrals whose upper limits are $a' = \min(a, \alpha)$ and $b' = \min(b, \beta)$ respectively. We obtain

$$\begin{aligned} \mu_2'(Y/A) = {} & \frac{4}{a^2b^2}\int_0^{a'}\int_0^{b'}\left\{1 + \frac{(\alpha-\xi)(\beta-\eta)-2\alpha\beta}{(a+\alpha)(b+\beta)}\right\}^k (a-\xi)(b-\eta)\,d\xi\,d\eta \\ & + \frac{4}{a^2b^2}\left\{1 - \frac{2\alpha\beta}{(a+\alpha)(b+\beta)}\right\}^k \left\{\int_0^a\int_0^b (a-\xi)(b-\eta)\,d\xi\,d\eta - \int_0^{a'}\int_0^{b'}(a-\xi)(b-\eta)\,d\xi\,d\eta\right\}. \end{aligned} \tag{40}$$

* The writer is indebted to Neyman & Bronowski for this convenient notation (see below).

By a simple change of variable we obtain

$$\mu_2'(Y/A) = \frac{4}{a^2b^2}\int_{[a-\alpha]}^{\alpha}\int_{[\beta-b]}^{\beta}\left\{1+\frac{uv-2\alpha\beta}{(a+\alpha)(b+\beta)}\right\}^k (u+a-\alpha)(v+b-\beta)\,du\,dv$$
$$+\frac{1}{a^2b^2}\left\{1-\frac{2\alpha\beta}{(a+\alpha)(b+\beta)}\right\}^k \{b^2[a-\alpha]^2+a^2[b-\beta]^2-[a-\alpha]^2[b-\beta]^2\}, \quad (41)$$

which is the result obtained by Bronowski & Neyman by a rather different method.*

8. Overlap of circles on a fixed circle

We now consider a fixed circle A of unit area and therefore of radius $b = 1/\sqrt{\pi}$, with k circles C of radius a dropped at random with their centres uniformly distributed over a circle T of radius $a+b$. The 2nd moment of the fraction Y/A of A not covered is, as in § 5, the expectation of $q^k(r)$, where $q(r)$ is the fraction of T not covered by two circles with centres falling at random in A a distance r apart. We have

$$q(r) = \frac{1+2a\sqrt{\pi}-\pi a^2+\Omega(r)}{1+2a\sqrt{\pi}+\pi a^2}, \quad (42)$$

where $\Omega(r)$, the overlap of two circles radius a with centres apart, is given by (19), (20) and (21) as before. We thus need the frequency distribution $\phi(r)$ of the distance between two points chosen at random in the circle of unit area to obtain

$$\mu_2'(Y/A) = \int_0^{2\sqrt{\pi}} q^k(r)\,\phi(r)\,dr. \quad (43)$$

To do this we use a fairly straightforward geometrical method, finding first the probability integral

$$F(r) = \int_0^r \phi(r)\,dr, \quad (44)$$

which is the probability that the distance between the two random points is less than r.

The probability that the first point is between v and $v+dv$ from the centre is $2v\,dv/b^2$, while if

$$A(v) = \text{area common to circles radii } b \text{ and } r \text{ with centres distance } v \text{ apart}, \quad (45)$$

it follows that the probability of the second point being within r of the first is $A(v)/\pi b^2$. Hence

$$F(r) = \int_0^b \frac{2v}{b^2}\frac{A(v)}{\pi b^2}\,dv. \quad (46)$$

Construct the triangle with sides r, b and v, and let the angles opposite to these be θ, ϕ and ψ. Then the following can be readily verified:

$$\text{If } r<b, \qquad \left.\begin{aligned} A(v) &= b^2\theta+r^2\phi-br\sin\psi &&\text{if } b-r<v<b,\\ &= \pi r^2 &&\text{if } 0<v<b-r.\end{aligned}\right\} \quad (47)$$

$$\text{If } b<r<2b, \qquad \left.\begin{aligned} A(v) &= b^2\theta+r^2\phi-br\sin\psi &&\text{if } r-b<v<b,\\ &= \pi b^2 &&\text{if } 0<v<r-b.\end{aligned}\right\} \quad (48)$$

The integration in (46) is carried out by parts, with ψ as the ultimate variable of integration, and to do this we obtain the result

$$A'(v) = -\frac{2br\sin\psi}{v}. \quad (49)$$

Putting

$$r = 2b\sin\tfrac{1}{2}\alpha, \quad (50)$$

we obtain, over the whole range of r,

$$F(r) = \frac{\alpha}{\pi}+r^2(\pi-\alpha)-\frac{r\cos\frac{1}{2}\alpha}{\sqrt{\pi}}-\frac{r^2\sin\alpha}{2}, \quad (51)$$

* And by Robbins (1945).

while differentiation yields the frequency distribution as

$$\phi(r) = 2r(\pi-\alpha) - \frac{\sqrt{\pi}}{2\cos\frac{1}{2}\alpha}(4r^2-\pi r^4). \tag{52}$$

It will be noted as a matter of interest that the chance of the two random points falling further apart than the radius of the circle is $1-F(b) = \frac{3\sqrt{3}}{4\pi}$ or 9/22 nearly.*

We thus obtain the 2nd moment of the uncovered area from (42), (43) and (52).

9. Use of probability generating functions

(i) *Binomial*

It is interesting to apply first the binomial distribution of k, the number of C's dropped uniformly and at random on T. Assume that T contains the centres of all the C's which touch or cover A, and that S is some larger area including T. If l C's are dropped at random on S, the probability generating function of the number of centres falling on T is

$$G_1(u) \equiv \left(\frac{S-T+Tu}{S}\right)^l. \tag{53}$$

Thus, from the general result of § 4, the mth moment of the fraction of A uncovered is the expectation of $\left(\frac{S-T+Tq}{S}\right)^l$, where q is the fraction of T uncovered by m C's falling on A. But the expression within brackets is the fraction of S uncovered. The use of the binomial generating function is thus verified.

(ii) *Poisson*

The Poisson distribution next suggests itself. If the number of centres follows this distribution with a mean of λ per unit area, the probability generating function is

$$G_2(u) \equiv e^{\lambda T(u-1)} \tag{54}$$

and the mth moment will be expectation over A of

$$e^{\lambda T(q-1)}.$$

Alternatively, we could write this as

$$\mu'_m(Y/A) = \exp(e^{-\lambda Z}), \tag{55}$$

where Z = area of overlap of m C's falling on A. In particular,

$$\text{mean value of } Y/A = \mu'_1(Y/A) = e^{-\lambda C}. \tag{56}$$

Thus the mth moment of Y/A is related to the characteristic function of Z, but this result does not appear to be of any theoretical importance: it does not, for instance, throw any light on the frequency distribution of Y/A. Formula (55) does, however, demonstrate the fact, which is otherwise obvious, that the area T does not enter into the frequency distribution of the fraction of A not covered by C's whose fall follows the Poisson distribution.

As far as the calculation of the variance is concerned, we need to calculate first the 2nd moment of $e^{-\lambda Z}$. In the cases where the falling areas are circles, the area of overlap Z of two circles is equal to $2C-\Omega(r)$, where $\Omega(r)$ is the function given above ((19) etc.) for the area common to two C's with centres r apart. The 2nd moment is thus

$$e^{-2\lambda C}\int e^{\lambda\Omega(r)}\phi(r)\,dr, \tag{57}$$

where $\phi(r)$ is the frequency function of r, the formulae for which are given above for the various cases.

* The solution to this problem (no. 698), given by Whitworth (1897), contains an error, resulting in the incorrect value of 35/88 nearly.

In the case of rectangles falling on rectangles, the necessary formula for the variance is given by Neyman & Bronowski in the form of a series, viz.

$$\mu_2(Y/A) = \frac{4e^{-2\alpha\beta\lambda}}{a^2b^2} \sum_{s=1}^{\infty} \frac{(\lambda\alpha\beta)^s \alpha\beta}{s!\,(s+1)^2\,(s+2)^2}$$
$$\times \{(s+2)\,a-\alpha+[\alpha-a]\,(1-a/\alpha)^{s+1}\}\{(s+2)\,b-\beta+[\beta-b]\,(1-b/\beta)^{s+1}\}. \tag{58}$$

Table 1. *Variance of fraction of fixed area not covered by areas C falling according to Poisson distribution*

(i) Circles falling on square; (ii) circles falling on rectangle 2×1; (iii) circles falling on rectangle 4×1; (iv) circles falling on circle; (v) squares falling on square (sides parallel).

Size of falling area C ÷ size of fixed area	Case	Mean area not covered		
		0·25	0·50	0·75
		Variance of area not covered		
0·2	(i)	0·0186	0·0303	0·0254
	(ii)	0·0182	0·0296	0·0248
	(iii)	0·0169	0·0275	0·0229
	(iv)	0·0190	0·0310	0·0260
	(v)	0·0182	0·0300	0·0252
1·0	(i)	0·0608	0·0964	0·0789
	(ii)	0·0573	0·0904	0·0743
	(iii)	0·0489	0·0766	0·0627
	(iv)	0·0620	0·0983	0·0810
	(v)	0·0593	0·0945	0·0781
1·8	(i)	0·0816	0·1262	0·1026
	(ii)	0·0771	0·1193	0·1040
	(iii)	0·0657	0·1015	0·0824
	(iv)	0·0829	0·1281	0·1040
	(v)	0·0790	0·1230	0·1002

In general, the variance increases in the following order as between the different combinations of shapes:

(1) Circles on rectangle 4×1, (iii).
(2) Circles on rectangle 2×1, (ii).
(3) Squares on square, (v).
(4) Circles on square, (i).
(5) Circles on circle, (iv).

There is an exception in the last case considered, however (mean fraction of area not covered $= 0{\cdot}75$, size of falling area ÷ fixed area $= 1{\cdot}8$), when the order is changed somewhat. However, the variances are generally of the same approximate magnitude.

(iii) *Contagious distribution*

Neyman & Bronowski have included in their study the case of a contagious law of type A with two parameters (see Neyman, 1939). Here the probability generating function is

$$G_3(u) = e^{m\{e^{\lambda T(u-1)}-1\}}. \tag{59}$$

They have pointed out that the expression of this as a series enables calculations to be utilized from the Poisson distribution.

In the general case the sth moment of the fraction of A not covered is the expectation of

$$e^{m(e^{-\lambda Z}-1)}, \tag{60}$$

where Z is the area common to s C's falling with their centres on A.

10. NUMERICAL RESULTS

It is impossible to calculate complete tables covering all cases, but it is of interest to calculate a few values for the purpose of comparison, and the following combinations of areas have been considered:

	Fixed area	Falling areas
(i)	Square	Circles
(ii)	Rectangle 2×1	Circles
(iii)	Rectangle 4×1	Circles
(iv)	Circle	Circles
(v)	Square	Squares (with sides parallel to fixed square)

In each case the fixed area has been made of unit size, while the falling areas were respectively $C = 0{\cdot}2$, $1{\cdot}0$ and $1{\cdot}8$. The areas were assumed to fall according to the Poisson distribution, the number of centres per unit area being such that the expected fraction of the fixed area A not covered was respectively 0·25, 0·5 and 0·75. Since from equation (56) the expected fraction of A not covered is $m = e^{-\lambda C}$, the relations between λ and C for the 9 combinations of m and C are

$$\lambda C = \log_e 4, \quad \log_e 2 \quad \text{and} \quad \log_e 4/3.$$

The 2nd moments were determined by quadrature from formula (57), where the falling areas were circles, and by direct evaluation of the series (58) for the case of squares on squares. The results are given in Table 1.

11. EXPERIMENTAL INVESTIGATION

Before the work of Robbins and Neyman & Bronowski was brought to the notice of the writer, an attempt was made to obtain experimentally a general formula for the variance of the fraction of area not covered. Attention was confined to the case of circles falling on squares, the centres of the former being chosen randomly (by means of random numbers) within the area T whose boundary is at a distance a outside the sides of the unit square A.

For each combination of C and k, samples of up to 200 in size were drawn. The various combinations were as given in Table 2.

Table 2. *Ranges of k and C covered in experimental determination of variance*

$\dfrac{\text{Area of circles dropped}}{\text{Area of square}} = C$	No. of circles $= k$
0·0077	5, 10, 15
0·031	5, 10, 15, 20, 30
0·033	5, 20, 40, 80, 120
0·25	1, 2, 4, 6
1	1, 2, 4, 6

Three methods were used to measure the fraction of the fixed square not covered in each sample. Method P involved the measurement of the covered area by planimeter. Method L utilized a photoelectric cell to measure the amount of light passing through a glass plate on which black paper disks had been stuck. Method C consisted of a simple count of squares on graduated paper, and generally this was the most convenient to operate. (The two neighbouring values of C were used to compare methods L and P.)

For each combination of C and k the average fraction not covered, $\bar{Y}$, was compared with the theoretical value

$$m = (1 - C/T)^k. \tag{61}$$

The observed standard deviation of the observations being s, the appropriate criterion for testing the mean is

$$t = \frac{\bar{Y}-m}{s/\sqrt{P}},$$

P being the number of observations in the sample. The results are given in Table 3.

Table 3. *Comparison between observed and expected values of fraction of area not covered*

P = planimeter method. L = photoelectric method. C = counting method.

Area of circle / Area of square $= C$	No. of circles k	No. in sample P	Mean area not covered		Standard deviation s	Deviation $t = \frac{\bar{Y}-m}{s/\sqrt{P}}$	Method of measurement
			Observed $\bar{Y}$	Expected m			
0·0077	5	100	0·9694	0·9683	0·0055	2·0000*	P
0·0077	10	100	0·9394	0·9376	0·0088	2·0454*	P
0·0077	15	100	0·9105	0·9079	0·0090	2·8889†	P
0·031	5	100	0·8924	0·8961	0·0222	−1·6667	P
0·031	10	200	0·8007	0·8030	0·0334	−0·9739	P
0·031	15	100	0·7134	0·7196	0·0404	−1·5347	P
0·031	20	100	0·6349	0·6449	0·0423	−2·3641*	P
0·031	30	100	0·5057	0·5179	0·0493	−2·4746*	P
0·033	5	100	0·8853	0·8901	0·0250	−5·9200†	L
0·033	20	100	0·6326	0·6278	0·0500	0·9600	L
0·033	40	100	0·3969	0·3941	0·0367	0·7629	L
0·033	80	75	0·1617	0·1553	0·0358	1·4756	L and P
0·033	120	50	0·0657	0·0612	0·0255	1·2478	P
0·25	1	30	0·8916	0·8950	0·0824	−0·2260	C
0·25	2	110	0·7853	0·8009	0·1214	−1·3477	C
0·25	4	110	0·6263	0·6415	0·1481	−1·0764	C
0·25	6	110	0·4891	0·5138	0·1307	−1·9820*	C
1·0	1	20	0·7583	0·7653	0·2350	−0·1332	C
1·0	2	50	0·5890	0·5856	0·2346	0·1025	C
1·0	4	50	0·3861	0·3430	0·2334	1·3057	C
1·0	6	50	0·1686	0·2008	0·1630	−1·3968	C

* Between 5 and 1% levels. † Beyond 1% level.

An examination of the values of t shows that too many of them are outside the 5% significance levels, while in each set corresponding to one value of C the values are too frequently of the same sign. The worst deviation is for $C = 0{\cdot}033$ and $k = 5$, with $t = -5{\cdot}92$, but this only corresponds to a difference between the observed mean of 0·885 and the theoretical mean of 0·890. The test is thus very sensitive and the deviations are not serious, and they arise from imperfections in the technique which have not been investigated in detail.

As regards random errors of measurement as distinct from bias, it was not possible to carry out a systematic estimation of the contribution of this source to the total variation. A series of repeated measurements for the case $C = 0{\cdot}031$, $k = 30$, for which the observed mean was 0·51, showed that the individual measurements had a standard error of about 0·009. As the total standard deviation in this case was 0·049, the true estimate of the standard error (i.e. omitting the error of measurement) was $\sqrt{(0{\cdot}049^2 - 0{\cdot}009^2)} = 0{\cdot}048$, and for our purposes this difference is negligible. There is thus some evidence for assuming that this method of estimating the variance was satisfactory.

12. Derivation of empirical formula for the variance

The consideration that the theoretical variance σ^2 of the fraction uncovered must be small whenever the mean m of this fraction is near the limits of its range, zero or unity, suggests that we might try the relation

$$\sigma^2 \sim m(1-m)$$

Table 4. *Derivation of empirical formula*

$\frac{\text{Area of circles}}{\text{Area of square}} = C$	No. of circles k	Ratio T/C	Observed variance $= s^2$	$g = \frac{s^2}{m(1-m)}$	Mean values of g $= g(C)$	$(T/C)^{\frac{3}{4}} \times g(C)$	Empirical value of variance $= \frac{2\cdot3m(1-m)}{(T/C)^{\frac{3}{4}}} = \sigma_1^2$	Exact value of variance $= \sigma^2$	Percentage error in σ_1^2	Percentage error in s^2
0·0077	5	155·7	0·0000303	0·000973			0·0000363	—	—	—
0·0077	10	155·7	0·0000774	0·001296	0·00108	2·10	0·0000692	—	—	—
0·0077	15	155·7	0·0000810	0·000980			0·0000990	—	—	—
0·031	5	46·1	0·000493	0·00530			0·000684	—	—	—
0·031	10	46·1	0·00112	0·00704			0·00116	—	—	—
0·031	15	46·1	0·00163	0·00810	0·00759	2·38	0·00148	—	—	—
0·031	20	46·1	0·00179	0·00780			0·00168	—	—	—
0·031	30	46·1	0·00243	0·00972			0·00184	—	—	—
0·033	5	43·5	0·000625	0·00638			0·000785	—	—	—
0·033	20	43·5	0·00250	0·01071			0·00188	—	—	—
0·033	40	43·5	0·00135	0·00564	0·00876	2·51	0·00192	—	—	—
0·033	80	43·5	0·00128	0·00976			0·00105	—	—	—
0·033	120	43·5	0·000650	0·01132			0·000461	—	—	—
0·25	1	9·5	0·00679	0·0723			0·00736	0·00731	0·7	− 7·1
0·25	2	9·5	0·0147	0·0924	0·0820	2·41	0·01249	0·0123	2·0	19·5
0·25	4	9·5	0·0219	0·0954			0·0180	0·0176	2·2	24·4
0·25	6	9·5	0·0171	0·0681			0·0196	0·0185	5·9	− 7·6
1·0	1	4·3	0·0552	0·307			0·0471	0·0508	−7·3	8·7
1·0	2	4·3	0·0550	0·223	0·234	2·05	0·0636	0·0651	−2·3	−15·5
1·0	4	4·3	0·0545	0·241			0·0590	0·0540	9·3	0·9
1·0	6	4·3	0·0266	0·166			0·0420	0·0339	23·9	−21·5

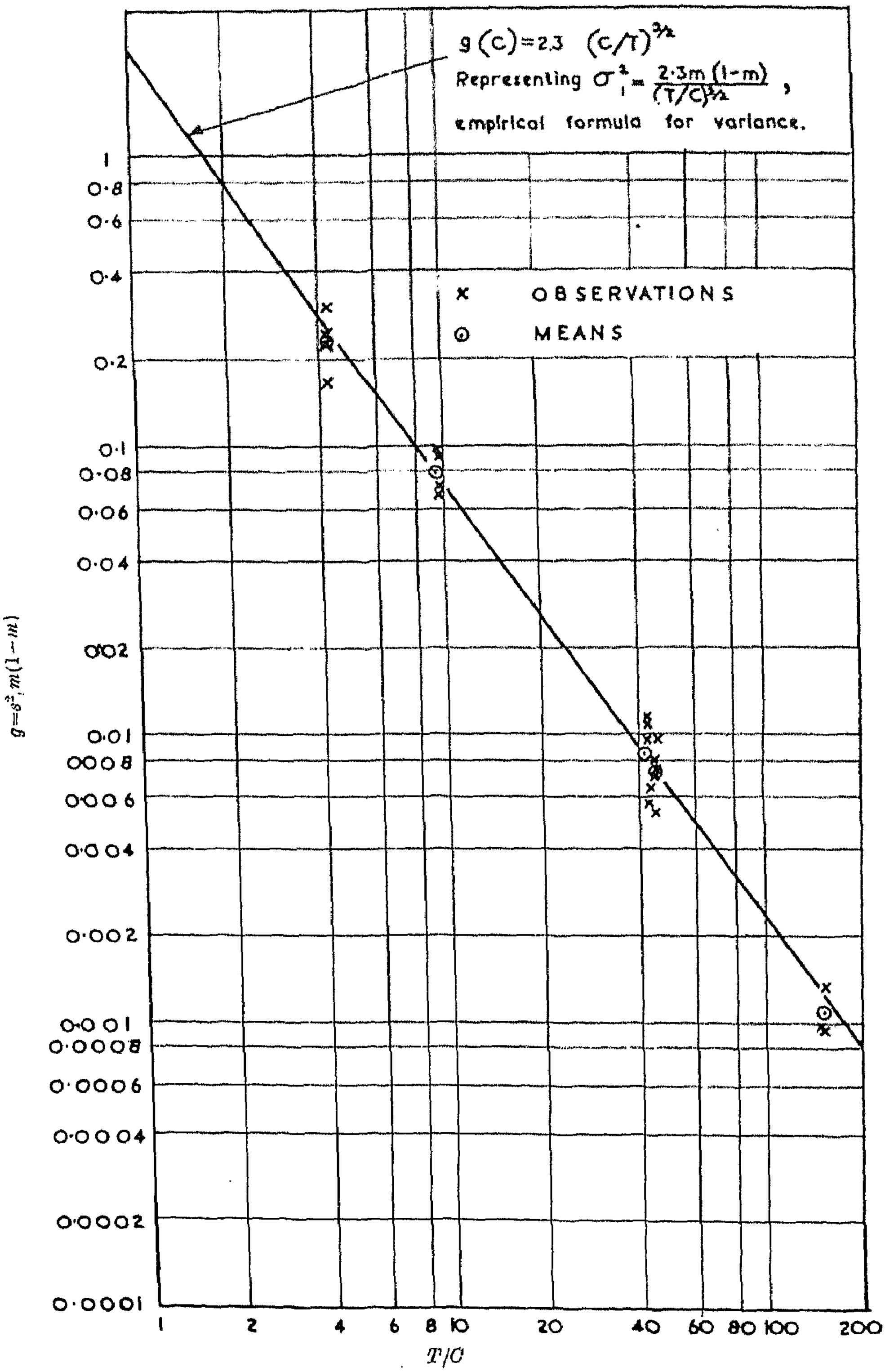


Fig. 2. Derivation of empirical formula for variance.

for given C. Accordingly we have calculated the quantity

$$g = \frac{s^2}{m(1-m)}, \tag{62}$$

and the results are given in Table 4.

For each value of C the values of g are by no means constant, but the variation is not excessive, and it is considered that for practical purposes we can take g to be a function of C, at least over the range considered. To obtain a suitable form for this function, it was decided, for very general reasons, to seek a simple relation between $g(C)$ and T/C, the latter being roughly the number of C's which could be placed on T if they could be fitted together without overlapping.

Table 5. *Comparison of empirical formula for variance with exact value*

σ^2 = exact value, σ_1^2 = empirical value, E = percentage error = $100\,(\sigma_1^2-\sigma^2)/\sigma^2$.

πa^2		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
0·2	σ^2	0·00501	0·00873	0·0114	0·0132	0·0144	0·0150
	σ_1^2	0·00516	0·00896	0·0117	0·0135	0·0147	0·0154
	E	+3·0	+2·6	+2·6	+2·3	+2·1	+2·7
0·4	σ^2	0·0156	0·0246	0·0290	0·0305	0·0300	0·0274
	σ_1^2	0·0149	0·0237	0·0284	0·0304	0·0305	0·0294
	E	−4·5	−3·7	−2·1	−0·3	+1·7	+7·3
0·6	σ^2	0·0277	0·0403	0·0447	0·0427	0·0395	0·0339
	σ_1^2	0·0257	0·0384	0·0431	0·0433	0·0407	0·0370
	E	−7·2	−4·7	−3·6	+1·4	+3·0	+9·1
0·8	σ^2	0·0398	0·0541	0·0552	0·0501	0·0428	0·0351
	σ_1^2	0·0365	0·0517	0·0551	0·0525	0·0471	0·0407
	E	−8·3	−4·4	−0·2	+4·8	+10·0	+16·0
1·0	σ^2	0·0508	0·0651	0·0628	0·0540	0·0436	0·0339
	σ_1^2	0·0471	0·0636	0·0648	0·0590	0·0507	0·0420
	E	−7·3	−2·3	+3·2	+9·3	+16·3	+23·9
1·2	σ^2	0·0605	0·0737	0·0677	0·0554	0·0427	0·0317
	σ_1^2	0·0571	0·0740	0·0724	0·0635	0·0525	0·0419
	E	−5·6	+0·4	+6·9	+14·6	+22·9	+32·5
1·4	σ^2	0·0689	0·0804	0·0706	0·0554	0·0410	0·0292
	σ_1^2	0·0667	0·0832	0·0786	0·0665	0·0531	0·0411
	E	−3·2	+3·5	+11·3	+20·0	+29·5	+40·7
1·6	σ^2	0·0763	0·0855	0·0723	0·0546	0·0389	0·0267
	σ_1^2	0·0757	0·0913	0·0834	0·0684	0·0530	0·0398
	E	−0·8	+6·8	+15·4	+25·3	+36·3	+49·1
1·8	σ^2	0·0828	0·0895	0·0730	0·0531	0·0367	0·0244
	σ_1^2	0·0843	0·0985	0·0873	0·0695	0·0524	0·0383
	E	+1·8	+10·1	+19·6	+30·9	+42·8	+57·0

Fig. 2 shows the result of plotting the observed mean value of $g(C)$ against T/C on logarithmic scales. The points (the values of $s^2/m(1-m)$ for various values of k are plotted in addition to the mean), lie reasonably close to a straight line of slope $-1{\cdot}5$, indicating the relation

$$g(C) \sim (C/T)^{\frac{3}{2}}. \tag{63}$$

The values of $(T/C)^{\frac{3}{2}}g(C)$ are shown in Table 4, where the values are seen to lie between 2 and 2·5 with an average of 2·3. Thus we derive the rough empirical formula

$$\left.\begin{aligned} \sigma_1^2 &= \frac{2{\cdot}3m(1-m)}{(T/C)^{\frac{3}{2}}}, \\ \text{where}\quad m &= (1-C/T). \end{aligned}\right\} \tag{64}$$

These are given in Table 4, together with the values in some cases of the percentage error of σ_1^2 compared with the true value σ^2 obtained from the methods described in § 5; the percentage error in the estimate s^2 observed experimentally is also given. (It was not possible to evaluate σ^2 in all cases, as the computation is somewhat laborious.) Another set of comparisons between the empirical and the exact formula is given below in Table 5, over the range $C = \pi a^2$ from 0·2 to 1·8 and $k = 1, 2, 3, 4, 5, 6$.

It will be seen that the empirical formula gives quite a satisfactory fit, e.g. with an error less than 10 %, over a considerable part of the range studied, but that the error tends to increase, i.e. the formula exaggerates the variance, as C and k increase.

13. Use of the empirical formula for the Poisson case

If k circles are dropped with their centres falling at random on the area T the mean area not covered can be written as

$$m(k) = (1 - C/T)^k,$$

and the empirical formula for the variance as

$$\mu_2(k) = \frac{2\cdot 3m(k)\,[1-m(k)]}{(T/C)^{\frac{1}{2}}}.$$

Table 6. *Comparison of empirical formula for variance of area not covered in case of circles falling on square according to Poisson distribution.*

Size of falling area C ÷ fixed area		Mean area not covered, m: 0·25	0·5	0·75
0·2	σ^2	0·0186	0·0303	0·0254
	σ_1^2	0·0196	0·0308	0·0257
	% error	5·4	1·7	1·2
1·0	σ^2	0·0608	0·0964	0·0789
	σ_1^2	0·0669	0·0981	0·0781
	% error	10·0	1·8	1·0
1·8	σ^2	0·0816	0·1262	0·1026
	σ_1^2	0·0942	0·1348	0·1057
	% error	15·4	6·8	3·0

Hence if k follows the Poisson distribution with expectation λT, the total variance based on the empirical formula is

$$\sigma_1^2 = \sum_{k=0}^{\infty} \frac{e^{-\lambda T}(\lambda T)^k}{k!}[m^2(k) + \mu_2(k)] - m^2,$$

where

$$m = \text{expected area not covered}$$
$$= e^{-\lambda C}. \tag{65}$$

We find after expanding $m^2(k)$ that

$$\sigma_1^2 = m^2\left[\left\{1 - \frac{2\cdot 3}{(T/C)^{\frac{1}{2}}}\right\} m^{-C/T} - 1\right] + \frac{2\cdot 3m}{(T/C)^{\frac{1}{2}}}. \tag{66}$$

This formula is compared with the exact values in Table 6 over the same range as in Table 1.

The agreement is again reasonably satisfactory over the greater part of the range, large positive errors occurring for large values of C and small values of m. These errors might be reduced by using a constant rather smaller than 2·3 in the empirical formula, but the point has not been investigated further.

SUMMARY

The mathematical study of bombing has given rise to the following problem. A fixed outline, such as a square or circle, is drawn on a plane, and other similar outlines are dropped at random on it. Estimates are then required of the variance of the fixed area which is not covered. Work by Robbins enables a theoretical formula to be derived, and Bronowski & Neyman have treated, by an independent method, the special case of rectangles falling on rectangles.

It is shown that in the case of circles falling on circles, squares or rectangles, the variance *can be expressed as the integral with respect to r of the product of two functions, one being* a simple function of the area of overlap of two circles with centres r apart, and the other being the frequency function of the distance r between two points chosen at random in the 'covered' area. This applies both to the case where the number of falling areas is fixed and where it follows a Poisson distribution. Numerical values have been calculated for a number of cases. An experimental method had been carried out prior to the above theoretical work, and the following empirical formula was derived for the variance of the fraction of a fixed square not covered by k circles, area C, falling at random on an area T containing centres of all C's which cover or touch the fixed square:

$$\sigma^2 = \frac{2\cdot 3m(1-m)}{(T/C)^{\frac{3}{2}}},$$

where m = mean fraction of area not covered $=(1-C/T)^k$.

This formula, and its extension to the Poisson case, have been shown to be in reasonable agreement with the exact values over a considerable range.

The writer is indebted to Miss G. O. Jeffcoate for valuable assistance in the computing and experimental work.

REFERENCES

BRONOWSKI, J. & NEYMAN, J. (1945). *Ann. Math. Statist.* **16**, 330.
NEYMAN, J. (1939). *Ann. Math. Statist.* **10**, 35.
ROBBINS, H. E. (1944). *Ann. Math. Statist.* **15**, 70.
ROBBINS, H. E. (1945). *Ann. Math. Statist.* **16**, 342.
WHITWORTH, W. A. (1897). *DCC Exercises in Choice and Chance.* Cambridge.

A STUDY OF A FIRST DYNASTY SERIES OF EGYPTIAN SKULLS FROM SAKKARA AND OF AN ELEVENTH DYNASTY SERIES FROM THEBES

By A. BATRAWI, Ph.D. and G. M. MORANT, D.Sc.

1. *Introduction.* This paper deals with forty-four male crania of 1st dynasty date (c. 3400 B.C.) discovered at Sakkara by Macramallah Effendi, who has published a report on the excavations (1940), and with fifty-five crania of 11th dynasty (c. 2000 B.C.) soldiers unearthed at Thebes in 1927 by the Egyptian Expedition of the Metropolitan Museum of Art, New York (Winlock, 1928). The cemetery at Sakkara, 20 miles south of Gizeh, was used by the middle classes of the local community. Prof. D. E. Derry has kindly provided the following notes on it:

The 1st dynasty cemetery at Sakkara excavated by Macramallah Effendi is of special interest. Comparatively few cemeteries of this date have been found, and, while the total number of forty-four skulls from which reliable measurements could be taken was small, yet the results yielded by these are such as to show that we are dealing with a race which differs in important features from those exhibited by the so-called predynastic people.

The observation that there were two races in Egypt in the early dynastic period was first made in the year 1909, when the results of measurements obtained from a series of male and female skulls of the 4th and 5th dynasties from the great necropolis surrounding the pyramids of Giza came to be examined and compared with crania from early predynastic graves. Until then the theory of an unbroken evolution of the Egyptian race from prehistoric times right through the dynastic period had been taught. It now became obvious that the culture which we know of as peculiarly Egyptian was associated with a race which could not have been derived from the predynastic people. The introduction of stone-working resulting in the erection of great tombs and statuary, as well as beautifully executed reliefs, paintings and above all writing, all pointed to a race far in advance of the predynastic people, who although skilled in the making of bowls and vases in stone as well as in pottery, and who had already attained to the discovery of the uses of copper, were, nevertheless, little removed from the Neolithic period.

The cemetery is unusual in consisting entirely of males. In the note on the skulls published in Macramallah Effendi's report it is stated that there were some females included in the collection. After the report had gone to press Macramallah Effendi informed the writer that a part of the cemetery was of 18th dynasty date. It turned out that all the female skulls came from this part and that therefore the 1st dynasty cemetery contained only remains of males. Dr Batrawi's examination of the figures confirms the statement made at the beginning of this note and shows the closeness of the relationship of the people of the 1st dynasty at Sakkara with those of the 4th and 5th dynasties from Deshasheh and Medum.

In his report on the discovery of the series of 11th dynasty skeletons at Thebes Mr H. E. Winlock (1928) says that they were found in 'a tomb in the row where the grandees of Mentuhotep's court had been buried'. He remarks:

Obviously what we had found was a soldiers' tomb. To judge from the cheapness of their burial they were only soldiers of the rank and file, and yet they had been given a catacomb presumably prepared for the dependents of the royal household, next to the tomb of the chancellor Khety. Clearly that was an especial honour. If we are right in supposing that all had been buried at once, they must have been slain in a single battle.

Prof. Derry examined the bodies on the spot, and he took measurements of the crania and of some of the long bones. About sixty bodies were counted and all proved to be those

of adult males who had died in the prime of life. Prof. Derry says that the skeletons were reburied after they had been examined.

2. *The measurements of the crania.* The Sakkara series was sexed and measured by Prof. Derry and we are indebted to him for allowing us to use his records in this paper. All the absolute measurements, given in Appendix II, are his readings with the exception of those of the *foramen magnum*, which were taken by one of the writers (A.B.) of this report. The measurements of the Thebes series were also kindly provided by Prof. Derry, together with means he had calculated. The readings for individual crania are given in Appendix III.

The technique of measurement followed by Prof. Derry is that of the Monaco Congress (Duckworth, 1913). He had used this when measuring the predynastic Egyptian series of skulls from Badari, of which part was remeasured later in London by Miss B. N. Stoessiger (1927), who followed the biometric technique. The two sets of measurements of the same fifty-three specimens have been compared (Morant, 1935), thus showing in detail what relations are to be expected between readings obtained by following the two techniques. These results were taken into account in preparing the definitions of Prof. Derry's measurements given in Appendix I below. The characters are denoted as far as possible there and in the tables by the customary index letters of the biometric technique.

3. *The nature of the two series.* Mean measurements and standard deviations for the two series are given in Table 1. The longest series of Egyptian skulls measured, known as the *E* series, came from a cemetery at Giza used from the 26th–30th dynasties (Davin & Pearson, 1924). Judging from comparisons of constants for a number of cranial characters, most of the other ancient Egyptian series described exhibit almost precisely the same order of variation as the one from Giza. In general they have been found to be rather less variable than European cranial series, while there is no evidence that there was any appreciable change in the variation exhibited by Egyptian populations during the long period from early predynastic to Roman times.

The two new series are shorter than several from Egypt previously described. Counting the number of characters for which the standard deviation for one series is greater or less than the corresponding constant for the other series, the situation is:

Sakkara and Thebes: Sakkara S.D. greater for nine and less for ten characters;

Sakkara and Giza: Sakkara S.D. greater for four and less for eleven characters;

Thebes and Giza: Thebes S.D. greater for eight and less for nine characters.

This crude comparison suggests that there can have been no marked differences between the variabilities of the three populations represented. As sets of differences are considered, the limit of significance accepted may be taken considerably higher than in the case of a single difference. Suppose that there is a real distinction if two of the standard deviations differ by an amount which is 3·5 or more times its probable error. Then one significant difference is found for the Sakkara and Thebes series (*NH*, *L*, Sakkara S.D. greater, $\Delta/\text{P.E.}\,\Delta = 3{\cdot}8$), none for the Sakkara and Giza, and three for the Thebes and Giza series (H' 4·1, S_1 3·5, S_2 3·5, Giza S.D. the greater in all three cases). The two new series are too short to give reliable comparisons, but the evidence suggests that the populations they represent were equally homogeneous, while both were rather less mixed in racial composition than the 26th–30th dynasty population of Giza.

4. *Comparisons of mean measurements.* Following biometric practice, it may be supposed in such a case that no statistical analysis of the series can reveal its racial components. The relationships of the series have to be judged by comparing them as wholes, on the basis of mean measurements, with other series known to exhibit unexceptional variation. It was shown by Morant (1925) that the recorded series of ancient Egyptian skulls can be divided into two groups. These were called, for convenience, the Upper and Lower Egyptian,

Table 1. *Means and standard deviations (with probable errors) of the Sakkara 1st dynasty and Thebes 11th dynasty series of male skulls*

Character*	Means		Standard deviations	
	Sakkara	Thebes	Sakkara	Thebes
L	186·9 ± 0·56 (41)	181·8 ± 0·53 (54)	5·31 ± 0·40	5·75 ± 0·37
B	138·7 ± 0·41 (43)	138·3 ± 0·41 (54)	3·99 ± 0·29	4·52 ± 0·29
B'	96·5 ± 0·39 (36)	93·6 ± 0·43 (55)	3·48 ± 0·28	4·72 ± 0·30
H'	135·4 ± 0·67 (32)	137·1 ± 0·37 (51)	5·63 ± 0·47	3·91 ± 0·26
[Aur. ht.]	114·8 ± 0·67 (27)	—	5·20 ± 0·48	—
LB	102·7 ± 0·57 (29)	100·7 ± 0·34 (46)	4·57 ± 0·40	3·40 ± 0·24
U	518·8 ± 1·5 (29)	507·4 ± 1·3 (50)	12·0 ± 1·1	13·2 ± 0·89
S_1	—	125·7 ± 0·47 (52)	—	5·01 ± 0·33
S_2	—	129·4 ± 0·56 (53)	—	6·00 ± 0·39
S_3	—	115·2 ± 0·82 (51)	—	8·63 ± 0·58
S	—	370·5 ± 1·2 (50)	—	12·7 ± 0·86
[Broca's Q']	—	300·3 ± 0·84 (49)	—	8·72 ± 0·59
fml	36·7 ± 0·26 (31)	—	2·18 ± 0·19	—
fmb	30·4 ± 0·22 (29)	—	1·74 ± 0·15	—
[$G'H$]	71·9 ± 0·55 (30)	72·0 ± 0·37 (45)	4·43 ± 0·39	3·71 ± 0·26
GB	96·5 ± 0·62 (25)	95·5 ± 0·48 (38)	4·57 ± 0·44	4·40 ± 0·34
J	127·8 (14)	127·6 ± 0·52 (32)	—	4·33 ± 0·36
[NH, L]	51·2 ± 0·50 (29)	51·8 ± 0·25 (45)	4·00 ± 0·35	2·52 ± 0·18
NB	25·4 ± 0·21 (30)	25·0 ± 0·20 (42)	1·70 ± 0·15	1·92 ± 0·14
[O_1']	38·9 ± 0·24 (26)	39·1 ± 0·16 (44)	1·84 ± 0·17	1·55 ± 0·11
[O_2]	32·5 ± 0·20 (26)	33·1 ± 0·23 (44)	1·50 ± 0·14	2·23 ± 0·16
[Prosthion GL]	99·6 ± 0·56 (26)	96·5 ± 0·47 (43)	4·23 ± 0·40	4·61 ± 0·34
100 B/L	74·2 ± 0·26 (39)	76·1 ± 0·26 (54)	2·44 ± 0·19	2·84 ± 0·18
100 H'/L	72·8 ± 0·41 (30)	75·5 ± 0·29 (51)	3·37 ± 0·29	3·06 ± 0·20
100 B/H'	102·6 ± 0·60 (31)	100·8 ± 0·41 (51)	4·95 ± 0·42	4·34 ± 0·29
100 fmb/fml	83·3 ± 0·72 (29)	—	5·76 ± 0·51	—
[100 $G'H/GB$]	74·3 ± 0·50 (25)	75·8 ± 0·53 (38)	3·70 ± 0·35	4·83 ± 0·37
[100 $NB/NH, L$]	49·5 ± 0·58 (29)	48·3 ± 0·48 (42)	4·63 ± 0·41	4·61 ± 0·34
[100 O_2/O_1']	83·6 ± 0·51 (26)	84·6 ± 0·49 (44)	3·86 ± 0·36	4·81 ± 0·35

* The characters are defined in Appendix I. A symbol in square brackets denotes either that the measurement is one not usually included in the biometric technique, or else that Prof. Derry's method of taking the measurements does not accord with biometric practice.

though there is evidence that the regions represented changed somewhat with time. The series in the first group came from the neighbourhood of Thebes and sites farther south, while those in the second group came from the same region of Upper Egypt and sites farther north. The first group includes all the predynastic series that have been described and some of dynastic date, the latest being of the 18th dynasty: the second group ranges from the 1st dynasty to Roman times, though no series available earlier than the 4th dynasty had come from the region immediately south of the Delta. The Sakkara series described in the present paper extends the range of such material back to the 1st dynasty.

It had been found that the means for all these series are almost constant for most of the metrical characters commonly recorded, but for a few measurements more significant differences are found and these separate the two groups of series. Characters of both kinds are treated in Table 2, which is based on Table XIII in Risdon's paper (1939) on the human remains from Lachish (Palestine). The first six characters are those which make the clearest distinction between the Upper and Lower Egyptian types of series, and they are all breadths or dependent on breadths—the latter being the horizontal circumference and the two indices—of the cranium. The Sakkara series is clearly assigned to the Lower Egyptian group, and if counted as a member of this the range of the mean minimum frontal breadths (B') for the group is slightly extended. The Thebes series is also assigned to the Lower Egyptian group by four of the six characters in question: for U and $100\,B/H'$, however, its means fall within the ranges given for the Upper Egyptian type of series.

Table 2. *Ranges of mean measurements for two groups of series of ancient Egyptian male crania and means for the Sakkara and Thebes series**

Series	Period	B	J	B'	U
Upper Egyptian type	Early predyn.–18th dyn.	131·4–134·3 (10)	123·6–127·5 (8)	90·4–92·8 (4)	500·0–510·4 (4)
Sakkara	1st dyn.	138·7	127·8	96·5	518·8
Thebes	11th dyn.	138·3	127·6	93·6	507·4
Lower Egyptian type	1st dyn.–Roman	135·3–139·3 (9)	127·5–131·3 (8)	93·0–96·2 (5)	510·8–518·7 (5)

Series	Period	$100\,B/L$	$100\,B/H'$	L	H'
Upper Egyptian type	Early predyn.–18th dyn.	71·7–73·7 (10)	98·1–101·1 (10)	182·2–185·2 (10)	132·4–135·9 (10)
Sakkara	1st dyn.	74·2	102·6	186·9	135·4
Thebes	11th dyn.	76·1	100·8	181·8	137·1
Lower Egyptian type	1st dyn.–Roman	73·7–76·0 (9)	102·3–106·4 (9)	181·4–185·8 (9)	130·7–136·0 (9)

* The characters are defined in Appendix 1. The numbers in brackets give the numbers of series to which the ranges relate. In the case of these previously described series the smallest number of crania on which any one of the means is based is 16, though this minimum number is about 30 for most of the characters. The numbers on which the Sakkara and Thebes means are based can be seen from Table 1, the only one less than 26 being 14 for the bizygomatic breadth (J) of the Sakkara series.

The last two characters in Table 2, which are the length and height of the cranium, fail to distinguish the two contrasted groups of series. The means for the two new series fall outside the ranges previously given by all the ancient Egyptian material, the Sakkara series giving the greatest L and the Thebes the greatest H'. The evidence of other characters must be taken into account, but so far the comparisons suggest that the two new series are of the Lower Egyptian type, and it is to be expected that they bear a closer resemblance to some of the series assigned to that group than to any other cranial series.

At the same time it may be noted that the Sakkara 1st dynasty and Thebes 11th dynasty populations are clearly differentiated by their mean cranial measurements. There are twenty characters in Table 1 for which means for both series are available. The most significant difference is for L, and it is 6·6 times its probable error, while five other characters—B', U,

prosthion GL, $100\,B/L$ and $100\,H'/L$—also show differences which exceed four times their probable errors.

5. *Comparisons by coefficients of racial likeness.* The method of Karl Pearson's coefficient of racial likeness has been applied extensively to series of ancient Egyptian crania. Risdon (1939) has given comparisons made in that way for twenty-two male series, including three from sites outside Egypt, and the treatment below is almost restricted to comparisons between these and the two new series described in the present paper. The procedure followed in applying the method described in several papers in *Biometrika* was adopted without modification.*

In deriving a classification of a number of cranial, or living, series from the coefficients of racial likeness found between them, it has been shown repeatedly that the most suggestive arrangement is obtained if the closest resemblances of the series, indicated by coefficients below a certain value, are alone taken into account. Risdon has given a diagram (1939, Fig. 3) showing all the reduced coefficients less than 5·0 between the twenty-two series with which he dealt. There are fifty-three of this lowest order among the 231 ($=22\times 21/2$) comparisons. The addition of the two new series to the classification referred to only requires a knowledge of the reduced coefficients less than 5·0 between them and the twenty-two series.

It has been pointed out that inspection of a few mean measurements can indicate whether a comparison of two particular series would almost certainly give a reduced coefficient greater than the limit chosen, or whether it might provide a value less than 5·0. The measurements used for this rapid test are six which are known to be those which show the most significant differences, and the greatest proportions of such differences, in comparisons of the group of series. These are the length, breadth and height of the brain-box and the three indices derived from these chords. For the fifty-three comparisons of the twenty-two

* A 'crude' coefficient is defined by

$$\frac{1}{m}S\left[\frac{n_s n_s'}{n_s+n_s'}\times\frac{(M_s-M_s')^2}{\sigma_s^2}\right]-1\pm 0\cdot 6745\sqrt{\frac{2}{m}},$$

where M_s is a mean based on n_s crania for the first series, M_s' and n_s' are the corresponding constants for the second series and m characters are compared. The σ's of the long 26th–30th dynasty Egyptian series were used throughout. The crude coefficient may be written

$$\frac{1}{m}S(\alpha)-1\pm 0\cdot 6745\sqrt{\frac{2}{m}},\quad \text{where}\quad \alpha=\frac{n_s n_s'}{n_s+n_s'}\times\frac{(M_s-M_s')^2}{\sigma_s^2}.$$

Its value is largely determined by the sizes of the two samples that happen to be available, if in fact they do not represent the same population. As many excavated crania are damaged to some extent, in the case of a particular series means for different characters will usually be based on various numbers of specimens (see Table 1). The mean number available for the characters used is denoted by $\bar{n}_s$ in the case of the first series and by $\bar{n}_s'$ in the case of the second series, and these 'sizes' of the samples are usually unequal and may be of very different orders. To obtain, as far as possible, a measure of the absolute divergence of the types compared which does not depend on the numbers of crania available, a 'reduced' coefficient of racial likeness is computed. This is defined to be

$$\frac{100\times 100}{100+100}\times\frac{\bar{n}_s+\bar{n}_s'}{\bar{n}_s\bar{n}_s'}\left[\frac{1}{m}S(\alpha)-1\pm 0\cdot 6745\sqrt{\frac{2}{m}}\right].$$

A reduced coefficient may be supposed a good approximation to the value which would be obtained if all the means for both series were for 100 individuals instead of for the numbers actually available. If a crude coefficient differs from zero by less than 3·5 times its probable error—a rare occurrence—then it is supposed that there is no evidence of a significant distinction between the two populations represented. In this case there is no need to compute a reduced coefficient. Otherwise, reduced coefficients are found and the classification of a number of series is based on these.

series giving reduced coefficients less than 5·0, the maximum differences for the six characters (in mm. or units of the indices) are:

L	B	H'	$100\ B/L$	$100\ H'/L$	$100\ B/H'$
3·1	3·0	3·5	2·0	2·3	2·8

To avoid the danger of missing comparisons which might be of the order required, in applying the test each of these values was increased arbitrarily by 0·2 giving:

L	B	H'	$100\ B/L$	$100\ H'/L$	$100\ B/H'$
3·3	3·2	3·7	2·2	2·5	3·0

In comparing a new series with the twenty-two it may be supposed that a reduced coefficient of racial likeness greater than 5·0 would almost certainly be found if the difference between the means is greater than the accepted limit in the case of any one or more of the six characters. For such comparisons the coefficients were not calculated. If the differences between the means are less than the limits for all six characters then a reduced coefficient less than 5·0 *might* be found: the coefficients were calculated in all such cases. In this way detailed comparisons were judged to be required between the new Sakkara, 1st dynasty, series, on the one hand, and six of the twenty-two treated by Risdon on the other; and between the new Thebes, 11th dynasty, on the one hand, and only two of the twenty-two series on the other. The previously described series involved in these two sets of comparisons—one series being included in both sets—are:

(i) Deshasheh and Medum, 4th and 5th dynasties (Thomson & MacIver, 1905). The two towns are south of Sakkara and both less than 40 miles from it.

(ii) Gizeh, 26th–30th dynasties (Davin & Pearson, 1924).

(iii) Sedment, 9th dynasty (Woo, 1930).

These three and the new Sakkara series are all from Lower Egypt among the total twenty-four series referred to above. All the other Egyptian sites mentioned are in Middle Egypt and close to Abydos and Thebes.

(iv) Abydos, 18th dynasty (Thomson & MacIver, 1905).

(v) Abydos, 1st dynasty, royal tombs (Morant, 1925).

(vi) Lachish, Palestine (Risdon, 1939). This series represents an Egyptian population. It is assigned to the seventh and eighth centuries B.C., though it is not well dated.

(vii) Tigré district, Abyssinia, modern (Sergi, 1912, means given in Morant, 1925).

(viii) Cretans, modern (von Luschan, 1913, means given in Woo, 1930). This series is not one of the twenty-two dealt with by Risdon. It was included because of its close resemblance to the new 11th dynasty series from Thebes. The test based on a comparison of the means of the six calvarial measurements shows that the only ancient Egyptian series which might give reduced coefficients less than 5·0 with the Cretan series are the Theban 11th dynasty and the Sedment series ((iii) above).

It must be emphasized that a reduced coefficient of racial likeness less than 5·0 represents a very close degree of resemblance. Values of that order have only been found between cranial series which would be expected, on account of their provenance, to represent the same or closely related populations. There is a danger that low reduced coefficients may be misleading owing to the influence on them of extraneous factors, such as inaccuracy in sexing or slight and unappreciated differences between the methods of measurement of two recorders working independently. It is safe to suppose that the two new series are made up entirely of the crania of adult males. In computing coefficients with them care

was taken to restrict a particular comparison to pairs of means based on measurements obtained by following precisely the same technique.

Owing partly to that restriction, the numbers of characters that could be used in computing coefficients with the new series are decidedly smaller than the 31 used ideally for the purpose. For these comparisons the smallest number of characters used is 9 and the largest number is 18.* Risdon (1939, pp. 131–2) has examined the matter experimentally and he concluded that use of a smaller number of characters—the set of 14 he considered being very similar to the sets we were able to use—can usually be expected to give a fairly close approximation to the result which would be obtained from about twice as many

Table 3. *Coefficients of racial likeness between ancient Egyptian, a Palestinian (Lachish) and modern series of male skulls from Abyssinia and Crete**

Series		Crude C.R.L. ± P.E.	Reduced C.R.L.
Sakkara, 1st dyn.	(32·1) with Deshasheh and Medum, 4th and 5th dyn. (46·0)	0·19 ± 0·32 (9)	—
	(32·1) with Abydos, 18th dyn. (49·9)	− 0·08 ± 0·32 (9)	—
	(31·6) with Lachish (249·3)	1·04 ± 0·25 (14)	1·85 ± 0·45
	(31·6) with Gizeh, 26th–30th dyn. (885·7)	2·45 ± 0·25 (14)	4·02 ± 0·41
	(31·6) with modern Abyssinian (61·4)	2·43 ± 0·25 (14)	5·82 ± 0·60
	(31·6) with Abydos, 1st dyn. royal tombs (33·6)	1·91 ± 0·25 (14)	5·87 ± 0·77
	(30·3) with Thebes, 11th dyn. (46·7)	4·26 ± 0·22 (18)	11·45 ± 0·59
Thebes, 11th dyn.	(49·2) with Sedment, 9th dyn. (37·9)	− 0·43 ± 0·28 (12)	—
	(49·0) with modern Cretans (50·4)	2·01 ± 0·30 (10)	4·04 ± 0·60
	(48·3) with Deshasheh and Medum, 4th and 5th dyn. (46·0)	2·23 ± 0·32 (9)	4·73 ± 0·68
Sedment, 9th dyn.	(37·5) with Deshasheh and Medum, 4th and 5th dyn. (39·9)	1·88 ± 0·25 (14)	4·86 ± 0·65
	(37·7) with modern Cretans (47·9)	2·71 ± 0·25 (15)	6·42 ± 0·59

* The numbers in brackets following the names of the series are the mean numbers of crania for the characters used in computing the coefficients. The numbers in brackets following the crude coefficients are the numbers of characters on which they are based. Woo (1930) gives coefficients with two of the series in the table above, and the values there differ from his because they were recalculated omitting the term $1/m$, which was discarded after 1930. The standard deviations of the long *E* series of 26th–30th dynasty crania from Gizeh (Davin & Pearson, 1924) were used in computing all the coefficients in the table.

characters. Occasionally, however, use of a smaller number of characters may suggest a rather misleading conclusion, and it will tend to indicate a rather wider separation of the series than that which would be found if all 31 characters could be used. With these reservations in mind the coefficients with the new series may be accepted as the best approximations it is possible to obtain in the circumstances.

All the coefficients of racial likeness found with the Sakkara 1st dynasty and Thebes 11th dynasty series are given in Table 3. Fig. 1 is a reproduction of part of a diagram given by Risdon (1939, p. 137) for the twenty-two ancient Egyptian and related series with which he dealt, with the addition of the two series described in the present paper and that of modern Cretans. The Sakkara series is seen to be an unexceptional member of the 'Lower Egyptian' constellation, having two insignificant coefficients and other close connexions

* The characters common to all the comparisons with the new series are L, B, H', LB, J, NB, $100\,B/L$, $100H'/L$ and $100B/H'$. Others used in some cases are B', U, S, fml, fmb and $100\,fmb/fml$, and for the coefficient between the two new series only $G'H$, NH, L, O_1', O_2, $100\,G'H/GB$, $100\,NB/NH$, L, $100\,O_2/O_1'$.

with members of that group. On the other hand, the 11th dynasty soldiers from Thebes clearly represent an Egyptian population of an aberrant type. The direct comparison fails to distinguish this from that of the 9th dynasty series from Sedment. Woo (1930) had found that the latter stands apart from all the other Egyptian series, and he pointed out that the Sedment bears a closer resemblance to a series of modern Cretans (von Luschan, 1913) than to most of the ancient Egyptian series. The Thebes 11th dynasty series has a reduced coefficient less than 5·0 with the Cretan, though the latter has no other coefficient of this order with any of the other Egyptian series.

6. *The racial history of ancient Egyptian populations.* The new evidence makes rather more precise the racial classification of ancient Egyptian populations given in earlier craniological papers in *Biometrika*. The 1st dynasty series of crania from Sakkara is the earliest in date that has been described representing the region immediately south of the Delta. It is an unexceptional representative of that group which must have prevailed

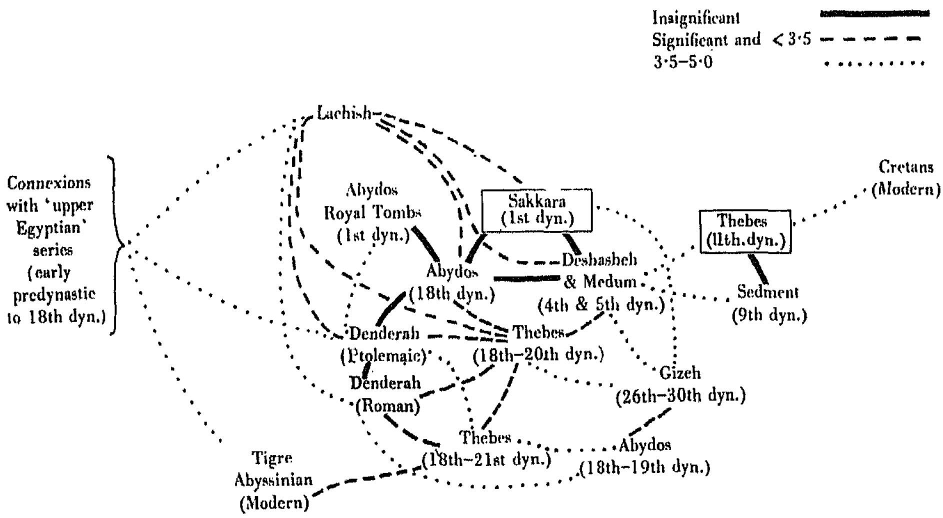


Fig. 1. Reduced coefficients of racial likeness between the two new and other ancient Egyptian and related series of male crania.

in the region—with only slight local and secular variants—from the 1st to the 30th dynasty and probably in both earlier and later periods as well. Such populations are said to be of 'Lower Egyptian' type.

In the region to the south, round Thebes and Abydos, the population was of a second racial type from the earliest predynastic (Badari) epoch for which there is any adequate craniological evidence. This is called the 'Upper Egyptian', though it would be better to call it Southern Egyptian. The population became modified slowly down to some time about the 18th dynasty. The change was such that the 'Upper Egyptian' type of population came to bear a closer and closer resemblance to the 'Lower Egyptian', though the two groups remained clearly distinct. About the 18th dynasty there must have been a fairly rapid, if not abrupt, change in the racial composition of the population of the Thebes and Abydos region. Nearly all the series from there, of that and later dates, are not of 'Upper' but of 'Lower Egyptian' type. They diverge slightly from the populations of the region immediately south of the Delta, however, in the direction of the 'Upper Egyptian' type. Six of these series—viz. those from Abydos, Thebes and Denderah of

dates ranging from the 18th dynasty to Roman times—are shown in Fig. 1. The 1st dynasty series from royal tombs at Abydos, also shown there, is an exception on account of its date. The obvious explanation of its peculiar position is that it represents an intrusive and more or less isolated community which was derived from the other centre of population to the north.

This accounts for twenty-two of the twenty-four series of crania considered. The classification of these does not seem to necessitate reference to any non-Egyptian peoples. This is not so, however, in the case of the remaining two series, viz. the new one of 11th dynasty soldiers from Thebes and the 9th dynasty series from Sedment (Deltaic region). These two might represent the same population as far as can be seen from the direct comparison, and both stand apart from the 'Lower Egyptian' constellation of series (see Fig. 1). The fact that the 11th dynasty series from Thebes has a close resemblance to one of Cretans, which is of modern date, suggests that the two aberrant communities in question may have been derived from the crossing of ancient Egyptians with people from some European or Asiatic source.

The mean basio-bregmatic heights (H'), cephalic indices and height-length indices are higher for the 11th dynasty Thebes and Sedment series than for any other of the series considered. The types, defined by average measurements, of these two thus diverge from that prevailing in ancient Egypt in the direction of the 'Armenoid' type. Elliot Smith (1911 and elsewhere) supposed that intrusive 'Armenoid' aliens played a considerable part in modifying the population of the country and that 'long before the time of the New Empire, Egypt was permeated from one end to the other with this foreign element'.

Our interpretation of the evidence fails entirely to support this hypothesis. There is no need to suppose that any people foreign to the country played a substantial part in modifying its population from predynastic to Roman times. The communities represented by the 11th dynasty Thebes and Sedment series may possibly have been derived from the crossing of Egyptian and 'Armenoid' people, but they stand apart. The remarkable point is not that two out of twenty-four populations should be peculiar in that way, but that the remaining twenty-two show interrelationships which do not suggest any admixture with alien stock. They can readily be explained on the supposition that there was a steady transference of population from the Deltaic region to the region of Thebes and Abydos, where the population was originally of a somewhat different type, from early predynastic times to the 18th dynasty. About that time the movement must have been accelerated, and thereafter the populations of the two centres were almost indistinguishable in racial type. The racial history of ancient Egypt was of a simple kind.

7. *Summary and conclusions.* This paper deals with forty-four male crania of 1st dynasty date from Sakkara and with fifty-five crania of 11th dynasty soldiers from Thebes. Individual measurements taken by Prof. D. E. Derry are given in appended tables. Judging from the rather small samples, the two populations represented exhibited the same order of variation, while both were rather less mixed in racial composition than the population of Giza from the 26th–30th dynasties. Mean measurements clearly differentiate the two new series from one another. Judging from characters considered singly, both series bear a close resemblance to some other ancient Egyptian series, and both are of 'Lower' rather than 'Upper Egyptian' type. Comparisons are made by the method of the coefficient of racial likeness, though decidedly fewer characters than the standard set of thirty-one used when possible are available for the purpose. The resulting relationships are shown in Fig. 1.

ARA*

Grave no.	$100 \frac{NB}{NH}, L$	$\left[100 \frac{O_2}{O_1'}\right]$	$100 \frac{fmb}{fml}$	Remarks
3	—	—	—	—
4	46·7	86·6	78·7	About 18 years. Metopic
7	50·4	81·3	79·2	Skull female? but pelvis definitely male
8	44·6	79·7	—	—
10	52·2	80·3	89·9	Root of nose depressed. Slight hydrocephaly?
14	45·2	78·0	82·9	—
16	53·2	94·4	87·5	About 16 years
17	—	—	81·8	—
18	—	—	—	—
21	—	—	78·4	—
23	—	—	—	—
27	—	—	93·1	—
30	53·9	80·0	88·9	—
32	50·4	85·0	77·5	—
38	46·9	82·3	92·6	—
49	53·7	83·1	79·5	—
50	52·1	90·0	84·1	About 19 years. Skull female? but pelvis definitely male
56	53·2	81·6	93·5	Distorted by grave pressure
58	50·5	79·0	82·1	—
59	49·1	—	—	—
62	53·2	79·0	88·6	About 18 years
65	44·3	—	—	—
72	40·7	86·8	80·6	—
77	—	—	—	Metopic
82	39·5	85·0	73·6	—
87	—	—	—	About 17 years
88	—	—	—	—
90	—	—	71·8	—
91	—	—	89·7	—
92	—	—	—	—
98	—	—	—	About 18 years
99	52·6	79·3	90·8	—
121	49·5	82·9	87·5	—
122	54·6	81·8	81·4	About 18 years
124	52·0	85·8	81·5	About 20 years
128	42·2	—	79·7	About 20 years
141	57·1	88·9	—	Metopic. Distorted
145	—	82·7	77·5	—
151	57·5	—	—	—
175	51·4	80·5	82·4	Old
180	46·4	88·2	83·8	—
184	47·2	87·2	—	—
190	47·1	85·5	78·4	About 17 years. Negroid
226	—	—	—	—

Derry's method of taking the measurement does not accord with biometric practice.

LDIERS FROM THEBES*

$100\frac{H'}{L}$	$100\frac{B}{H'}$	$\left[100\frac{G'H}{GB}\right]$	$\left[100\frac{NB}{NH}, L\right]$	$\left[100\frac{O_2}{O_1'}\right]$	Remarks
76·5	100·0	70·7	46·6	80·5	—
79·6	102·6	77·4	46·9	85·5	—
74·1	98·5	—	39·5	94·9	—
77·2	98·2	68·7	60·0	76·2	Metopic
76·6	97·5	78·4	43·4	86·8	Metopic
74·6	95·7	73·5	51·0	90·9	—
74·2	102·2	76·7	46·7	90·2?	—
72·5	96·5	71·1	49·5	80·0	—
71·7	104·1?	—	—	—	—
79·1	98·6	81·5?	46·2	90·5	—
79·1	88·9	82·8	45·5	84·3	Left parietal fractured
80·6	94·3	74·7	53·1	73·2	—
72·2	105·7	—	46·8	94·9	—
70·8	106·7	71·9	48·1	75·6	—
74·1	98·5	76·7	47·1	76·3	—
75·2	97·8	66·8	49·5	84·2	—
75·1	102·2	71·7	48·1	82·1	—
69·7	110·8	75·8	60·0	88·2	Metopic
81·7	98·6	88·3	40·0?	85·9	—
74·5	104·0	70·6	52·4	83·3	—
69·9	102·9	75·7	51·0	79·0	—
75·9	102·5	—	48·0	84·2?	Fractures in parietal, frontal and orbital regions
80·9	98·3	80·4	42·6	89·0	—
78·4	93·2	77·9	46·2	92·0	—
74·1	101·8	74·5	43·4	89·7	Metopic
75·5	101·4	78·7	45·4	82·5	Metopic
—	—	82·5	46·2	86·7	Metopic
77·5	100·4	69·5?	—	—	Metopic
77·2	94·2	79·5	47·4	93·8	—
76·7	107·6	80·1	53·1	87·0	Metopic
73·9	104·5	77·5	45·3	86·4	Metopic
—	—	—	—	—	Metopic
73·5	104·9	—	51·0	80·0	—
73·4	103·2	64·8	51·9	83·3	—
69·3	107·6	75·7	40·7	81·3	—
73·4	96·8	73·9	51·0	84·4	—
78·9	96·4	79·0	49·5	79·5	—
72·5	100·4	72·0	60·0	81·6	—
77·1	105·2	74·2	49·0	87·8	—
78·0	99·3	80·1	43·1	88·5	—
72·0	103·8	—	—	83·1	—
80·5	95·4	78·3	49·0	78·4	—
74·7	101·5	78·4	44·7	79·8	—
71·8	105·9	—	50·0	78·6	—
79·1	97·1	72·2	48·5	84·3	Metopic
71·9	109·1	—	—	—	Metopic
73·2	99·3	—	—	—	Metopic
—	—	78·2	50·5	85·4	Metopic
78·3	97·9	—	—	92·1	—
75·5	103·6	—	—	—	—
76·2?	99·3?	—	—	—	—
—	—	—	—	—	—
78·5	102·2	—	—	—	—
78·0	100·0	—	—	—	—
73·8	103·0	—	—	—	—

r else that Prof. Derry's method of taking the measurement does not accord with biometric practice.
:ceed fifty-nine.

The Sakkara 1st dynasty series, which is the earliest from the region immediately south of the Delta, is an unexceptional member of the 'Lower Egyptian' constellation, and it can be supposed to typify the population of Northern Egypt at the time. The 11th dynasty series of soldiers from Thebes is linked to the same group, but it diverges from it. The type is indistinguishable from that of a 9th dynasty series, from Sedment. The former also has a link with the type of a series of modern Cretans. The two aberrant communities of Thebes and Sedment must be supposed to have been derived from the crossing of ancient Egyptians with people from some European or Asiatic source. Our knowledge of the racial history of ancient Egypt derived from craniological evidence is reviewed.

REFERENCES

DAVIN, A. G. & PEARSON, K. (1924). On the biometric constants of the human skull. *Biometrika*, **16**, 328–63.
DUCKWORTH, W. L. H. (1913). *International Agreements for the Unification (a) of Craniometric and Cephalometric Measurements, (b) of Anthropometric Measurements to be made on the Living Subject.* Camb. Univ. Press.
LUSCHAN, F. VON (1913). Beiträge zur Anthropologie von Kreta. *Z. Ethn.* **14**, 307–93.
MACRAMALLAH, R. (1940). *Un Cimetière Archaïque de la Classe Moyenne à Saqquara.* Imp. Nat. le Caire.
MORANT, G. M. (1925). A study of Egyptian craniology from prehistoric to Roman times. *Biometrika*, **17**, 1–52.
MORANT, G. M (1935). A study of predynastic Egyptian skulls from Badari based on measurements taken by Miss B. N. Stoessiger and Prof. D. E. Derry. *Biometrika*, **27**, 293–309.
RISDON, D. L. (1939). A study of the cranial and other human remains from Palestine excavated at Tell Duweir (Lachish) by the Wellcome-Marston archaeological research expedition. *Biometrika*, **31**, 99–166.
SERGI, S. (1912). *Crania Habessinica. Contributo all' Antropologia dell' Africa Orientale.* Rome: Loescher.
SMITH, G. ELLIOT (1911). *The Ancient Egyptians and their Influence upon the Civilization of Europe.* Harper.
STOESSIGER, B. N. (1927). A study of the Badarian crania recently excavated by the British School of Archaeology in Egypt. *Biometrika*, **19**, 110–50.
THOMSON, A. & RANDALL-MACIVER, D. (1905). *The Ancient Races of the Thebaid.* Oxford Univ. Press.
WINLOCK, H. E. (1928). The Egyptian Expedition, 1925–27. *Bull. Met. Mus. Art, New York.*
WOO, T. L. (1930). A study of seventy-one ninth dynasty Egyptian skulls from Sedment. *Biometrika*, **22**, 65–93.

APPENDIX I. DEFINITIONS OF MEASUREMENTS

Individual measurements of the two series of crania are given in Appendices II–III. The contractions used there and in tables in the text to denote characters are:

L = maximum glabella-occipital length. B = maximum horizontal breadth. B' = minimum frontal breadth. H' = basio-bregmatic height. *Aur. ht.* = 'vertical height from line joining highest points of external auditory meatuses'. LB = basion to nasion. U = maximum horizontal circumference above the superciliary ridges. S_1 = arc nasion to bregma. S_2 = arc bregma to lambda. S_3 = arc lambda to opisthion. S = total sagittal arc from nasion to opisthion. *Broca's* Q' = transverse arc from 'the most prominent point on the posterior root of the left zygoma, exactly above the auditory aperture', to the same point on the right passing through the bregma. fml = basion to opisthion. fmb = maximum breadth of foramen magnum. $G'H$ = nasion to alveolar point. GB = facial breadth between lowest points on zygomatico-maxillary sutures. J = maximum breadth between zygomatic arches. NH, L = nasal height from nasion to point furthest removed from it on the margin of the left pyriform aperture. NB = maximum breadth of the pyriform aperture. O_1' = breadth of right orbit from the dacryon. O_2 = maximum height of right orbit. *Prosthion* GL = basion to prosthion.

THE GENERALIZATION OF 'STUDENT'S' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARIANCES ARE INVOLVED

By B. L. WELCH, B.A., Ph.D.

1. *Introduction and summary.* Let η be a population parameter which is estimated by an observed quantity y, normally distributed with variance σ_y^2. Let $\sigma_y^2 = \sum_{i=1}^{k} \lambda_i \sigma_i^2$, where the λ_i are known positive numbers and the σ_i^2 are unknown variances. Suppose that the observed data provide estimates s_i^2 of these variances, based on f_i degrees of freedom, respectively, so that the sampling distribution of s_i^2 is

$$p(s_i^2)\,ds_i^2 = \frac{1}{\Gamma(\frac{1}{2}f_i)}\left(\frac{f_i s_i^2}{2\sigma_i^2}\right)^{\frac{1}{2}f_i-1} \exp\left[-\frac{1}{2}\frac{f_i s_i^2}{\sigma_i^2}\right] d\left(\frac{f_i s_i^2}{2\sigma_i^2}\right), \tag{1}$$

and that these estimates are distributed independently of each other and of y.

A very simple particular case of this set-up occurs when we have samples of n_1 and n_2, respectively, from two normal populations with true means α_1 and α_2 and standard deviations σ_1 and σ_2. If η is the true difference $(\alpha_1 - \alpha_2)$ between the means, the estimated difference is $y = (\bar{x}_1 - \bar{x}_2)$. The variance of the estimate is $\sigma_y^2 = (\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)$, where $\lambda_1 = 1/n_1$ and $\lambda_2 = 1/n_2$. The estimated values of σ_1^2 and σ_2^2 are $s_1^2 = \Sigma_1/f_1$ and $s_2^2 = \Sigma_2/f_2$, where Σ_1 and Σ_2 are the respective sums of squares of observations from the individual sample means and $f_1 = (n_1 - 1)$ and $f_2 = (n_2 - 1)$. These s^2 are distributed in the form (1) and the postulated conditions of independence hold.

Another particular case, again with $k = 2$, arises when we wish to compare two regression coefficients, fitted to independent sets of data, without making the assumption that the population residual variance about the true regression line is the same for both sets.

The present paper is written mainly with these practical applications of the case $k = 2$ in mind, but the results are expressed generally for any k, since no further analytical difficulties are involved. It will be shown how probability statements about y, considered as an estimate of η, may be made similar in character to those which W. S. Gosset derived for the mean of a single sample of n observations ('Student', 1908). We shall, in effect, seek a quantity h, calculable from the observations, with the property that the chance of the difference $(y - \eta)$ falling short of h is a given probability P. It is clear that h must be a function of the individual variances s_i^2 and of P. If the abbreviation Pr. is used to mean 'the probability of the relation in the bracket following', our problem is to satisfy the equation

$$\Pr.\,[(y-\eta) < h(s_1^2, s_2^2, \ldots, s_k^2, P)] = P. \tag{2}$$

In Gosset's case the solution was, of course, simply

$$\Pr.\,[(\bar{x}-\alpha) < t_P s/\sqrt{n}] = P, \tag{3}$$

where t_P is the value, corresponding to the probability level P, in the 'Student' t-distribution with $f = (n-1)$ degrees of freedom.

In the next section the mathematical derivation of the exact solution of (2) is given. This is then followed by some consideration of its expression in numerical terms. First, a series solution in powers of $1/f_i$ is developed, which may be used for calculating tables. Then some comparisons are made with a non-series approximate solution which is based on a particular way of regarding the distribution of a quantity of the general form $z = (\Sigma a_i \chi_i^2)$.

Some brief discussion is then added which may serve to place the present contribution in its proper relationship to other papers which have been written on this topic.

Finally, it is shown how the inequality (2) may be adapted to provide an interval estimate for η.

2. *Mathematical derivation of solution.* Let $j(s_1^2, s_2^2, \ldots, s_k^2, P)$ denote the probability that $(y-\eta)$ is less than $h(s_1^2, s_2^2, \ldots, s_k^2, P)$, *given* s_i^2 $(i = 1, 2, \ldots, k)$. Then, since y is distributed quite independently of the estimated variances, we have

$$j(s_1^2, s_2^2, \ldots, s_k^2, P) = \int_{u=-\infty}^{\frac{h(s_1^2, s_2^2, \ldots, s_k^2, P)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2}\, du = I\left\{\frac{h(s_1^2, s_2^2, \ldots, s_k^2, P)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\}, \tag{4}$$

where I is used to denote the normal probability integral. The condition of equation (2) is then simply that, if $j(s_1^2, s_2^2, \ldots, s_k^2, P)$ is averaged over the probability distributions of s_i^2 as given by (1), the result will equal P. Thus

$$\int_{s^2} \cdots \int j(s_1^2, s_2^2, \ldots, s_k^2, P) \prod_i p(s_i^2)\, ds_i^2 = P. \tag{5}$$

Now we may expand $j(s_1^2, s_2^2, \ldots, s_k^2, P)$ about an origin $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$ in a Taylor expansion. Thus

$$j(s_1^2, s_2^2, \ldots, s_k^2, P) = \exp\left[\sum_i (s_i^2 - \sigma_i^2)\, \partial_i\right] j(w_1, w_2, \ldots, w_k, P), \tag{6}$$

it being understood that the exponential is to be expanded in a power series in ∂_i and that ∂_i is to be interpreted so that

$$\partial_i^r j(w_1, w_2, \ldots, w_k, P) = \left[\frac{\partial}{\partial w_i^r} j(w_1, w_2, \ldots, w_k, P)\right]_{w_j = \sigma_j^2}. \tag{7}$$

On making the substitution of (6) into (5) our result may be written

$$\Theta j(w_1, w_2, \ldots, w_k, P) = P, \tag{8}$$

where

$$\Theta = \prod_i \int \exp\left[(s_i^2 - \sigma_i^2)\partial_i\right] p(s_i^2)\, ds_i^2. \tag{9}$$

Now, substituting into (9) from (1), the integral comes out in simple form, i.e.

$$\begin{aligned}
\Theta &= \prod_i \left\{1 - \frac{2\sigma_i^2 \partial_i}{f_i}\right\}^{-\frac{1}{2}f_i} \exp\left[-\sigma_i^2 \partial_i\right] \\
&= \exp\left\{-\Sigma\sigma_i^2\partial_i - \tfrac{1}{2}\Sigma f_i \log\left(1 - \frac{2\sigma_i^2\partial_i}{f_i}\right)\right\} \\
&= \exp\left\{\Sigma\frac{\sigma_i^4\partial_i^2}{f_i} + \frac{4}{3}\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2} + 2\Sigma\frac{\sigma_i^8\partial_i^4}{f_i^3} + \text{etc.}\right\} \\
&= 1 + \Sigma\frac{\sigma_i^4\partial_i^2}{f_i} + \left\{\frac{4}{3}\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2} + \frac{1}{2}\left(\Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right)^2\right\} + \text{etc.}
\end{aligned} \tag{10}$$

Substituting (4) into (8) we have finally

$$\Theta I\left\{\frac{h(w_1, w_2, \ldots, w_k, P)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} = P. \tag{11}$$

This, in a very condensed form, is the solution to our problem.* The operator Θ constitutes a direction to carry out the partial differentiations indicated by (10). w_j must then be equated to σ_j^2. The solution of the resulting equation will give $h(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2, P)$ and therefore the required $h(s_1^2, s_2^2, \ldots, s_k^2, P)$.

3. *The development of the series solution.* It will be convenient to write $h(w)$ for $h(w_1, w_2, \ldots, w_k, P)$ and ξ for the normal deviate such that $I(\xi) = P$. We may then expand $I\left\{\frac{h(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\}$ in a Taylor series about ξ as origin. Thus

$$I\left\{\frac{h(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} = \exp\left[\left\{\frac{h(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} - \xi\right\} D\right] I(v), \tag{12}$$

it being understood that the exponential is to be expanded in powers of D, and that these powers are to be interpreted so that

$$D^r I(v) = \left[\frac{d^r}{dv^r} I(v)\right]_{v=\xi}. \tag{13}$$

Equation (11) then becomes

$$\Theta \exp\left[\left\{\frac{h(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} - \xi\right\} D\right] I(v) = I(\xi). \tag{14}$$

This may now be solved by successive approximations.

The initial approximation is the large-sample normal approximation

$$h_0(w) = \xi\sqrt{(\Sigma\lambda_i w_i)}, \tag{15}$$

and we may write

$$h(w) = \xi\sqrt{(\Sigma\lambda_i w_i)} + h_1(w) + h_2(w) + \text{etc.}, \tag{16}$$

where $h_1(w)$ includes terms of order $1/f_i$, $h_2(w)$ terms of order $1/f_i^2$ and so on. For the moment we shall treat terms of the order $1/f_i^3$ as negligible. Then (14) gives

$$\Theta \exp\left[\left\{\frac{\xi\sqrt{(\Sigma\lambda_i w_i)}}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} - \xi\right\} D\right] \exp\left[\left\{\frac{h_1(w) + h_2(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\} D\right] I(v) = I(\xi) \tag{17}$$

i.e. $$\Theta \exp\left[\xi D\left\{\sqrt{\frac{\Sigma\lambda_i w_i}{\Sigma\lambda_i\sigma_i^2}} - 1\right\}\right]\left[1 + \frac{h_1(w)\,D}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} + \left\{\frac{h_2(w)\,D}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} + \frac{1}{2}\frac{h_1^2(w)\,D^2}{\Sigma\lambda_i\sigma_i^2}\right\} \ldots\right] I(v) = I(\xi). \tag{18}$$

Or, using (10),

$$\begin{aligned}
&\left[\frac{h_1(\sigma^2)\,D}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} + \Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\exp\left(\xi D\left\{\sqrt{\frac{\Sigma\lambda_i w_i}{\Sigma\lambda_i\sigma_i^2}} - 1\right\}\right)\right] I(v) \\
&\quad + \left[\frac{h_2(\sigma^2)\,D}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}} + \frac{1}{2}\frac{h_1^2(\sigma^2)\,D^2}{\Sigma\lambda_i\sigma_i^2} + \Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\exp\left(\xi D\left\{\sqrt{\frac{\Sigma\lambda_i w_i}{\Sigma\lambda_i\sigma_i^2}} - 1\right\}\right)\frac{h_1(w)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right. \\
&\quad \left. + \left\{\frac{4}{3}\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2} + \frac{1}{2}\left(\Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right)^2\right\}\exp\left(\xi D\left\{\sqrt{\frac{\Sigma\lambda_i w_i}{\Sigma\lambda_i\sigma_i^2}} - 1\right\}\right)\right] I(v) = 0.
\end{aligned} \tag{19}$$

The equation of the first order term to zero gives

$$h_1(\sigma^2) = \frac{\xi(1+\xi^2)}{4}\frac{\left(\Sigma\frac{\lambda_i^2\sigma_i^4}{f_i}\right)}{(\Sigma\lambda_i\sigma_i^2)^{\frac{3}{2}}}. \tag{20}$$

* Equation (11) can also be expressed as an integral equation and this form may be necessary for providing numerical values where the f_i are very small.

This can then be substituted in the second-order term which, when equated to zero, will give $h_2(\sigma^2)$. The process may obviously be extended to higher orders, although the expressions become so complex that a slightly different procedure has then been found to be preferable. To terms of order $1/f_i^2$ our solution is

$$h(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}\left[1+\frac{(1+\xi^2)}{4}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2}-\frac{(1+\xi^2)}{2}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^2}\right.$$
$$\left.+\frac{(3+5\xi^2+\xi^4)}{3}\frac{\left(\Sigma\frac{\lambda_i^3 s_i^6}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^3}-\frac{(15+32\xi^2+9\xi^4)}{32}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4}\right]. \tag{21}$$

It may be noted that in the particular case $k = 1$, this reduces, as it should, to the already known expansion of the deviate of the straightforward 'Student' distribution (Fisher, 1941, p. 151), viz.

$$t_P = \xi\left[1+\frac{(1+\xi^2)}{4f}+\frac{(3+16\xi^2+5\xi^4)}{96f^2}+\text{etc.}\right]. \tag{22}$$

It is proposed in another communication to give tables of $h(s^2)$ based on the expansion (21) carried to some further terms.

4. *Discussion of a non-series approximation.* It will be recalled that in Gosset's original approach to the single sample problem ('Student', 1908) his initial step was to note that the first four moments of the distribution of s^2 were consistent with the assumption that the distribution could be represented by a Pearson Type III curve. He was fortunate in this way to rediscover a distribution which had already been found by Helmert, as this permitted him to go on to the derivation of the t-distribution. In our present case, as in many others arising naturally in statistical work, we are led to consider, instead of s^2, a linear function $\Sigma\lambda_i s_i^2$ of several s_i^2. If this linear function were distributed in a Pearson Type III distribution a whole range of new problems could be dealt with by well-established theory. However, in general, we do not have this good fortune. For $\Sigma\lambda_i s_i^2$ is of the form $\Sigma a_i\chi_i^2$, where $a_i = \lambda_i\sigma_i^2/f_i$, and the distribution of this quantity is only of Type III if all the a_i, except one, are zero, or if all the a_i happen to be equal.

Nevertheless, for practical purposes an *approximation* to the distribution of $\Sigma\lambda_i s_i^2$, using a Type III curve with start, mean and variance suitably adjusted, can still be useful. In two previous papers (Welch, 1936, 1938) I have employed this method to obtain numerical comparisons of the merits of different statistical procedures, where full calculations with the true distributions would have been unduly laborious. The method of determining the constants in the approximation was given for the case $k = 2$ in the first of these papers and is as follows.

If $z = (a\chi_1^2+b\chi_2^2)$, and the approximate distribution curve is written in the form

$$p(z)\,dz = \frac{1}{\Gamma(\frac{1}{2}f)}e^{-\frac{1}{2}(z/g)}\left(\frac{z}{2g}\right)^{\frac{1}{2}f-1}d\left(\frac{z}{2g}\right), \tag{23}$$

then making the first two moments of (23) agree with the true moments of z, we find

$$f = \frac{(af_1+bf_2)^2}{a^2f_1+b^2f_2},\quad g = \frac{a^2f_1+b^2f_2}{af_1+bf_2}. \tag{24}$$

Phrasing the matter rather differently, we can say that z/g is approximately distributed as

χ^2 with degrees of freedom f. Of course f, given by (24), will in general be fractional, but the letter used to designate this quantity was chosen, and the term 'effective degrees of freedom' has been used, because by doing so we can appeal immediately to a considerable body of further theoretical results.

In particular we can say that the criterion

$$v = \frac{(y-\eta)}{\sqrt{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}} \tag{25}$$

follows approximately the 'Student' t-distribution with degrees of freedom

$$f = \frac{(\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)^2}{\dfrac{\lambda_1^2\sigma_1^4}{f_1} + \dfrac{\lambda_2^2\sigma_2^4}{f_2}}. \tag{26}$$

More generally, when k is not restricted to 2, the same line of argument leads us to say that the criterion

$$v = \frac{(y-\eta)}{\sqrt{(\Sigma\lambda_i s_i^2)}} \tag{27}$$

is approximately distributed as 'Student's' t with degrees of freedom

$$f = \frac{(\Sigma\lambda_i\sigma_i^2)^2}{\Sigma\dfrac{\lambda_i^2\sigma_i^4}{f_i}}. \tag{28}$$

Not knowing the σ_i's in (28), there are several ways in which we may now proceed, depending on what weight we may be willing to attach to any vague *a priori* notions we may possess of their *relative* magnitudes (cf. Welch, 1938). If we are not willing to assume anything, perhaps the best choice is

$$f = \frac{(\Sigma\lambda_i s_i^2)^2 - 2\left(\Sigma\dfrac{\lambda_i^2 s_i^4}{f_i+2}\right)}{\left(\Sigma\dfrac{\lambda_i^2 s_i^4}{f_i+2}\right)}. \tag{29}$$

It may be shown that the numerator of (29) has, in repeated samples, an average value $(\Sigma\lambda_i\sigma_i^2)^2$, and the denominator has average value $\Sigma\lambda_i^2\sigma_i^4/f_i$. In a certain sense, therefore, (29) is a fair estimate of (28).

To sum up, then, the interpretation of y as an estimate of η, using the present type of approximation involves only the reference of the criterion (27) to tables of the 'Student' distribution, entered with degrees of freedom given by (29).

Some further light is now thrown on this procedure by the expansion for the exact solution of our problem derived in the preceding section. For the implications of referring v to the 'Student' distribution may be seen by substituting f from (29) into the expansion (22) of the 'Student' deviate. On doing this and then expanding in powers of $1/f_i$ it is found that, in effect, our approximation corresponds to assuming that

$$h(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}\left[1 + \frac{(1+\xi^2)}{4}\frac{\left(\Sigma\dfrac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2} - \frac{(1+\xi^2)}{2}\frac{\left(\Sigma\dfrac{\lambda_i^2 s_i^4}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^2} + \frac{(51+64\xi^2+5\xi^4)}{96}\frac{\left(\Sigma\dfrac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4} + \dots\right], \tag{30}$$

whereas, in fact, the true solution is given by (21). Comparison shows that we have exact

agreement to terms of order $1/f_i$ and in the first of the quadratic terms. To the second order the difference between the expressions in square brackets in equations (21) and (30) is

$$\frac{(3+5\xi^2+\xi^4)}{3}\left\{\frac{\left(\Sigma\frac{\lambda_i^3 s_i^6}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^3}-\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4}\right\}. \tag{31}$$

This difference vanishes if any one of the s_i^2 is overwhelmingly larger than all the others, or if s_i^2 is proportional to f_i/λ_i. It appears that, in general, the difference is not likely to be large. We have, therefore, found some justification for using the Type III approximation in the present case.

The above comparison has been made on the basis of the series developments, but it should be borne in mind that approximations based on positive frequency functions, such as those falling under the Pearson system, usually provide a higher degree of accuracy than might appear from any consideration of expansions. Furthermore, they are apt to give an insight into the nature of the situation which may sometimes be lost in working out the details of exact solutions. In the present case I feel that the comparison of this section serves to give added confidence in the exact solution,* which I have put forward in the previous two sections, quite as much as it demonstrates the value of the approximate method.

5. *Further discussion.* In comparing the present contribution with other work on the subject, the essential point to notice is the averaging process involved in equation (5). We are not trying here to make probability statements valid for *fixed* s_i^2, but are averaging over the joint probability distribution of the s_i^2, taking into account, therefore, the different values which can arise by chance in sampling from populations with fixed σ_i^2.

This averaging over the joint distribution of the s_i^2 is parallel to the step taken in Section III of Gosset's original memoir (1908) where, in effect, he starts with the distribution of t for samples with *fixed* s and then averages over the distribution of s which he has already derived earlier. He thus arrives at the unrestricted distribution of t (or, more strictly, of a quantity z, which is equal to t multiplied by a constant). This distribution forms the basis of the significance tests which he illustrates in his Section IX and of the method of deriving interval estimates for the population mean which he outlines in his Section VIII.

In the present paper the parallelism with Gosset's work may be obscured to some extent by the fact that we do not from the outset seek the probability distribution of some pivotal quantity like t, explicitly expressed. It so happens that we are able to proceed to a method of deriving an expansion for the required probability level without making explicit reference to such a quantity. Nevertheless there remains the important resemblance with Gosset's development, in that we do not confine ourselves to samples with fixed s_i^2.

This procedure stands in sharp contrast to the formulation of the problem of comparing two means, favoured by R. A. Fisher (e.g. 1941) and H. Jeffreys (1940). These writers prefer a solution which they ascribe initially to W. U. Behrens (1929). Looked at from one point of view, Behrens's paper appears to contain some gross algebraical errors. Fisher and Jeffreys, however, develop lines of argument by means of which they claim that Behrens's solution is quite justified. It seems to me difficult to say how far (if at all) any of these arguments may have been in Behrens's mind when he wrote his paper and I shall not attempt to elucidate this question here. We may, however, permit ourselves one observation about the developments according to Fisher and Jeffreys.

* Exact in the sense that it is independent of the irrelevant population parameters σ_i^2.

Both these writers, at some stage, limit the field of their probability inferences to a subset in which the s_i^2 are regarded as *fixed*. In order to solve the problem on these lines Jeffreys introduces an *a priori* distribution function for the unknown σ_i, following his general philosophy for dealing with such questions. Fisher, on the other hand, arrives at the same answer by a special utilization of what he terms the *fiducial* distribution of σ_i.

Jeffreys's approach here does not raise any new issues to those who are familiar with the general body of his researches on statistical inference. Fisher's justification of Behrens's solution is perhaps of more immediate interest as it raises controversial points which are important more specifically in relation to our present topic of discussion. For although Fisher's approach has been very much criticized by a number of writers, starting with M. S. Bartlett (1936), the critics have not wished to throw doubt on the whole body of results which Fisher includes under the heading of fiducial inference. The criticism has been for the most part selective, directed mainly at the way in which so-called *simultaneous* fiducial distributions of several parameters have been defined and manipulated.

I have, myself, quite definite views on these questions (particularly on the usage of the word 'fiducial') but do not feel that I need express them at any great length here. I disagree with Fisher, but this divergence of opinion must already have become apparent in the way I have defined the field within which I make my probability inferences about η. It appears to me to be quite artificial to restrict our view to one which, even in a limited sense, fixes s_i^2. It is true that, in the two-sample problem, we have to draw our inferences from the unique pair of samples observed, or, more precisely, from the statistics $\bar{x}_1$, $\bar{x}_2$, s_1^2 and s_2^2 which they provide. These statistics are our only *data* for the purpose of making inferences, but we add something to these data in the *interpretation* when we regard the samples as being drawn randomly from hypothetical normal populations. Once having embarked on this method of interpretation, we should stick to it consistently throughout. The sampling variations of s_i^2 should be taken into account only by a direct use of the probability distributions as given by our equation (1) and not by any inversion such as is involved in Fisher's conception of the fiducial distribution of σ_i^2. As we have seen, it is quite possible to make probability statements about the difference between the population means without making any reference whatever either to inverse probability or to fiducial distributions.

The distinction between the procedure which Fisher advocates and one which averages over the s^2 distributions has, of course, been stressed by most of the writers who have contributed papers on the subject, from whatever viewpoint (e.g. Bartlett, 1936, p. 566, and Yates, 1939.) What has been lacking hitherto, however, is a solution, analogous to Gosset's single sample solution, which makes complete use of the information contained in the data provided. Bartlett indicated one particular way in which probability inferences about the difference between two population means might be made, but was careful to point out that the problem of making the best possible inferences (in the theoretical sense of utilizing all the information in the data to its full extent) was still an open one. There has indeed been some doubt expressed whether a fully satisfactory solution from this point of view existed at all. I believe, however, that the one I advance above in equation (11), and develop in equation (21), meets all the requirements that one can reasonably expect.

Whatever conclusion the reader may come to on these matters, however, he will probably wish to know how, in the numerical details, this solution will differ from that of Behrens. This will be more easily seen when some tables become available, but fortunately certain

comparisons can already be made. For Fisher (1941, p. 155) has provided a series expansion of the Behrens solution. In our notation, and with $k=2$, this may be written, to order $1/f_i$, as follows:

$$h(s^2) = \xi\sqrt{(\lambda_1 s_1^2+\lambda_2 s_2^2)}\left[1+\frac{(1+\xi^2)}{4}\frac{\left(\frac{\lambda_1^2 s_1^4}{f_1}+\frac{\lambda_2^2 s_2^4}{f_2}\right)}{(\lambda_1 s_1^2+\lambda_2 s_2^2)^2}+\left(\frac{1}{f_1}+\frac{1}{f_2}\right)\frac{\lambda_1\lambda_2 s_1^2 s_2^2}{(\lambda_1 s_1^2+\lambda_2 s_2^2)^2}\right]. \qquad (32)$$

Even to this order, this differs from our equation (21) in the inclusion of an extra term. In other words, although the two solutions are the same when samples are large enough to adopt the large-sample normal approximation, they differ immediately we take into account the first corrective term, i.e. they differ as soon as we begin to attach any importance to 'Studentization'.

6. *An interval estimate for* η. We have shown in §§2 and 3 how to calculate a value $h(s_1^2, s_2^2, \ldots, s_k^2, P)$, depending on the observed variances $s_1^2, s_2^2, \ldots, s_k^2$, such that the probability is P that $(y-\eta) < h(s_1^2, s_2^2, \ldots, s_k^2, P)$. This provides a method of testing the consistency of an observed y with a prescribed value η.

When the question is not whether any particular given η is contradicted by the data, but rather one of estimating η and at the same time of providing a measure of the uncertainty of the estimate, the further step required is immediate. For, as in the case of a single sample, the order of the words in our probability statement can be changed so that it becomes—the probability is P that η is greater than $\{y-h(s_1^2, s_2^2, s_k^2, P)\}$. An interval estimate for η is then obtained by taking two levels P_1 and P_2 for P. Thus the probability is (P_1-P_2) that η lies between $\{y-h(s_1^2, s_2^2, \ldots, s_k^2, P_1)\}$ and $\{y-h(s_1^2, s_2^2, \ldots, s_k^2, P_2)\}$.

If $P_2 = (1-P_1)$ the range will be symmetrically placed about y. Thus, for example, if $P_1 = 0{\cdot}95$ and $P_2 = 0{\cdot}05$, the chance will be 90 % that η lies within the range

$$y \pm 1{\cdot}6449\sqrt{(\Sigma\lambda_i s_i^2)}\left[1+\frac{1+(1{\cdot}6449)^2}{4}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2}+\text{etc.}\right]. \qquad (33)$$

It may be noted, incidentally, that this range is always narrower than similar ranges calculated from Behrens's solution.

REFERENCES

BARTLETT, M. S. (1936). The information available in small samples. *Proc. Camb. Phil. Soc.* **32**, 560–6.

BEHRENS, W. U. (1929). Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen. *Landw. Jb.* **68**, 807–37.

FISHER, R. A. (1941). The asymptotic approach to Behrens's integral, with further tables for the *d* test of significance. *Ann. Eugen., Lond.*, **11**, 141–72.

JEFFREYS, H. (1940). Note on the Behrens-Fisher formula. *Ann. Eugen., Lond.*, **10**, 48–51.

'STUDENT' (1908). The probable error of a mean. *Biometrika*, **6**, 1–25.

WELCH, B. L. (1936). Specification of rules for rejecting too variable a product, with particular reference to an electric lamp problem. *J. Roy. Statist. Soc. Suppl.* **3**, 29–48.

WELCH, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, **29**, 350–62.

YATES, F. (1939). An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters. *Proc. Camb. Phil. Soc.* **35**, 579–91.

THE DISTRIBUTION OF KENDALL'S τ COEFFICIENT OF RANK CORRELATION IN RANKINGS CONTAINING TIES

By G. P. SILLITTO, Ph.D., B.Sc., *Research Department, I.C.I. (Explosives Ltd.)*

A new coefficient of rank correlation has been described by Kendall (1938, 1942, 1943) and denoted by him as τ. This coefficient has advantages over Spearman's ρ in respect of the smoothness of its distribution and the rapidity with which it approaches normality, thus facilitating significance testing, and in being readily adapted to cases of partial rank correlation.

The distribution of τ has been worked out by Kendall (1938, 1943) for cases in which neither ranking contains members which are graded equal, i.e. rankings containing no 'ties'. It is the purpose of the present paper to deal with cases, which frequently arise in practice, in which ties occur in one of the two rankings. The method is a generalization of that of Kendall and will be given in some detail for the case of tied pairs, while the results of further generalization to multiplet ties will be indicated without detailed proof, which can in all cases be effected simply on the lines indicated.

Definition of τ for rankings containing ties

In counting the 'score' of a pair of rankings, by the methods suggested by Kendall, each member is compared with the other members of the same ranking, and additions to or subtractions from the score are made depending on whether it is smaller or greater in each case. If some members are ranked equal then it is proposed that no change be made in the score in comparing them. This obviously accords with the intuitive aspects of ranking. Thus in the pair of rankings following, the score is $+8$:

$$\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 3 \end{array}$$

The maximum score possible is thus obviously reduced by the presence of ties, and it is evident that the presence of each tied pair reduces the maximum possible score by unity, so that it becomes $\frac{1}{2}n(n-1)-p_2$ for the case of a ranking of n members containing p_2 pairs. Thus for such a ranking τ would be defined as

$$\tau = 2S/\{n(n-1)-2p_2\},$$

where S is the observed score.

Generally, each r-tuplet tie reduces the maximum possible score by $\frac{1}{2}r(r-1)$* so that for a ranking of n members containing p_2 pairs, p_3 triplets, ..., p_r r-tuplets,

$$\tau = \frac{2S}{n(n-1)-2p_2-6p_3-\ldots-r(r-1)p_r}.$$

The sum of the frequencies of the possible scores

When no ties are present, each permutation of the n members produces a possible score so that there are in all $n!$ possible scores. When ties are present they decrease the number of possible permutations of an assigned set of members, but, on the other hand, they give rise

* This result has been given by Kendall (1945).

to further families of scores due to the different places in the ranking which can be occupied by the tied members. Thus, for instance, the rankings

$$113456, \quad 122456, \quad 123356, \quad 123446, \quad 123455$$

all give rise to the maximum score, 14.

Considering any assigned ranking, the number of possible permutations with p_2 pairs present is $n!/2^{p_2}$, or if there are in addition p_3 triplets, ..., p_r r-tuplets, $n!/(2!)^{p_2}(3!)^{p_3}\ldots(r!)^{p_r}$.

The distribution of scores of an assigned set of ranks will be referred to as the basic distribution for the type of ranking concerned, since consideration of the possible ways of assigning the p_2 pairs, p_3 triplets, etc., among the members of the ranking has only the effect of multiplying the frequency of each score by a constant factor. This factor is the number of ways of distributing the $p_1+p_2+p_3+\ldots+p_r$ ranks among the n members. This is the number of possible permutations

$$\frac{(p_1+p_2+p_3+\ldots+p_r)!}{p_1!\,p_2!\,p_3!\ldots p_r!}.$$

BASIC FREQUENCY DISTRIBUTIONS OF THE SCORES

The basic frequency distributions can be established by an extension of the methods given by Kendall. Considering first the case of tied pairs the frequency function of the basic distribution of the scores may be written $f(S, n, p_2)$, where p_2 is the number of pairs. The frequency generating function is then $\sum_j f(S_j, n, p_2)\,t^{S_j}$. Now consider the addition of another tied pair, with a greater ranking than any of the existing ranks. If it is added to the extreme left of the ranking it adds $-2n$ to the score. Moving one of the pair one place to the right adds 2 to this new score; bringing the other added member up to it adds another 2. Starting again with both the new members on the extreme left, movement of one of them two places to the right adds 4 to the new score, bringing the other up to it in two steps each of one place, adds successively a further 2 and 4. Proceeding in this way all possible additions to the old score which may be brought about by the addition of a tied pair of new members are represented by the array

$$\begin{array}{cccccc}
-2n & -(2n-2) & -(2n-4) & -(2n-6) & \ldots & 0 \\
 & -(2n-4) & -(2n-6) & -(2n-8) & & +2 \\
 & & -(2n-8) & -(2n-10) & & +4 \\
 & & & -(2n-12) & & \ldots \\
 & & & & & \ldots \\
 & & & & & +2n
\end{array}$$

Thus the addition of a new tied pair has the effect of multiplying the frequency generating function by

$$\{t^{-2n} + (t^{-(2n-2)} + t^{-(2n-4)}) + (t^{-(2n-4)} + t^{-(2n-6)} + t^{-(2n-8)}) + \ldots + (t^0 + t^2 + \ldots + t^{2n})\}.$$

The addition of a single new member to the ranking has the effect of multiplying the frequency generating function by

$$\{t^{-n} + t^{-(n-2)} + t^{-(n-4)} + \ldots + t^n\}$$

as shown by Kendall, the presence of tied pairs in the existing ranking having no effect.

With these two recurrence relations there is no difficulty in drawing up a table of basic frequency distributions for tied pairs as exemplified in Table 1, in which only positive values of the score are shown, negative values being obtainable by symmetry.

Table 1. *Distribution of the score S for values of n from 3 to 7, and for rankings containing p_2 pairs of members ranked equal (only positive half of symmetrical distribution)*

n	p_2	Values of S																					
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
3	.	.	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	1	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	.	6	.	5	.	3	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	1	.	3	.	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	2	2	.	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
5	.	22	.	20	.	15	.	9	.	4	.	1	.	.	.	.	.	.	.	.	.	.	.
5	1	.	11	.	9	.	6	.	3	.	1	.	.	.	.	.	.	.	.	.	.	.	.
5	2	6	.	5	.	4	.	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
6	.	.	101	.	90	.	71	.	49	.	29	.	14	.	5	.	1	.	.	.	.	.	.
6	1	52	.	49	.	41	.	30	.	19	.	10	.	4	.	1	.	.	.	.	.	.	.
6	2	.	26	.	23	.	18	.	12	.	7	.	3	.	1	.	.	.	.	.	.	.	.
6	3	14	.	12	.	11	.	7	.	5	.	2	.	1	.	.	.	.	.	.	.	.	.
7	.	.	573	.	531	.	455	.	359	.	259	.	169	.	98	.	49	.	20	.	6	.	1
7	1	292	.	281	.	250	.	205	.	154	.	105	.	64	.	34	.	15	.	5	.	1	.
7	2	.	146	.	135	.	115	.	90	.	64	.	41	.	23	.	11	.	4	.	1	.	.
7	3	74	.	72	.	63	.	52	.	38	.	26	.	15	.	8	.	3	.	1	.	.	.

Before the construction of the table has proceeded far, however, it becomes evident that there is a recurrence relation between individual frequencies for any given value of n, such that the frequency of any score S_j for p_2 pairs is the sum of the frequencies of $S_j - 1$ and $S_j + 1$ for $p_2 + 1$ pairs. This obviously arises from the fact that if two members ranked equal, say rth, in a ranking with $p_2 + 1$ pairs are subsequently distinguished and given rankings r and $r+1$, this will increase the score by unity if the $(r+1)$ member falls after the (r) member when the ranking is arrayed against another ranking in the natural order $1, 2, 3, \ldots, n$, and reduce it by unity if the other member of the pair becomes the $(r+1)$th; and these two possibilities complete the ways of forming a ranking with p_2 pairs from one with $p_2 + 1$ pairs.

This simple relationship, which may be written

$$f(S_j, n, p_2) = f(S_j + 1, n, p_2 + 1) + f(S_j - 1, n, p_2 + 1),$$

or taking another way of writing the basic distribution function

$$\phi(S_j, p_1, p_2) = \phi(S_j + 1, p_1 - 2, p_2 + 1) + \phi(S_j - 1, p_1 - 2, p_2 + 1), \tag{1}$$

p_1 being the number of members not in tied pairs, is of great assistance in tabulating the frequency distribution, and will be used below to establish the formula for the variance of S. It can be generalized to cover the effect of increasing the number of r-tuplets, when it becomes

$$\begin{aligned} \phi(S_j, p_1, p_2, \ldots, p_{r-1}, p_r) = {} & \phi(S_j - \overline{r-1}, p_1 - 1, p_2, \ldots, p_{r-1} - 1, p_r + 1) \\ & + \phi(S_j - \overline{r-3}, p_1 - 1, p_2, \ldots, p_{r-1} - 1, p_r + 1) \\ & + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & + \phi(S_j + \overline{r-1}, p_1 - 1, p_2, \ldots, p_{r-1} - 1, p_r + 1). \end{aligned} \tag{2}$$

FREQUENCY AND PROBABILITY DISTRIBUTIONS OF THE SCORE S

From a table of basic frequency distributions such as Table 1 the construction of a table showing the probability of attaining or exceeding an observed value of S by chance from an uncorrelated pair of rankings can obviously be constructed, and Table 2 shows such probabilities (positive S only, negative values obtainable by symmetry) for values of n up to 10, and all possible numbers p_2 of paired and p_3 of triplet ties.

THE VARIANCE OF S

The variance of S when ties are present can be readily derived by using the recurrence relations given above and the value given by Kendall for the case of no ties. For the case of tied pairs consider

$$\begin{aligned}(S+1)^2\phi(S+1, p_1-2, p_2+1)+(S-1)^2\phi(S-1, p_1-2, p_2+1)\\ = S^2\{\phi(S+1, p_1-2, p_2+1)+\phi(S-1, p_1-2, p_2+1)\}\\ +2\{(S+1)\phi(S+1, p_1-2, p_2+1)-(S-1)\phi(S-1, p_1-2, p_2+1)\}\\ -\{\phi(S+1, p_1-2, p_2+1)+\phi(S-1, p_1-2, p_2+1)\}.\end{aligned}$$

If now both sides of this equation are summed over all values of S, the terms on the left-hand side become

$$\frac{n!}{(2!)^{p_2+1}}\operatorname{var}\phi(S, p_1-2, p_2+1).$$

The first terms on the right-hand side become by virtue of the recurrence relation (1)

$$\frac{n!}{(2!)^{p_2}}\operatorname{var}\phi(S, p_1, p_2),$$

the second vanishes through the symmetry of the distribution, while the third becomes

$$-\frac{2\,.\,n!}{(2!)^{p_2+1}}.$$

Hence there is obtained

$$\operatorname{var}\phi(S, p_1-2, p_2+1) = \operatorname{var}\phi(S, p_1, p_2)-1,$$

and so

$$\begin{aligned}\operatorname{var}\phi(S, n-2p_2, p_2) &= \operatorname{var}\phi(S, n)-p_2\\ &= \frac{n(n-1)(2n+5)}{18}-p_2,\end{aligned}$$

using Kendall's result.

These results can also be generalized, using equation (2), to deal with multiplet ties, obtaining

$$\operatorname{var}\phi(S, p_1-1, \ldots, p_{r-1}-1, p_r+1) = \operatorname{var}\phi(S, p_1, \ldots, p_{r-1}, p_r)-(r^2-1)/3,$$

and

$$\operatorname{var}\phi(S, p_1, p_2, \ldots, p_r) = \frac{n(n-1)(2n+5)}{18}-p_2-\frac{3+8}{3}p_3-\ldots-\frac{3+8+\ldots+(r^2-1)}{3}p_r.$$

It is obvious from these equations for multiplet ties that for any given number of ties of each multiplicity the variance will tend towards that of the system without any ties as n increases.

APPLICATION TO A PRACTICAL CASE

The following results were obtained in a practical case in which two different tests were carried out on one each of a set of products. The problem is to determine the degree of relationship between the results of the two tests. It is also an instance of an occurrence

which arises at times in practice, in which some of the results are 'off the scale' of measurement with respect to one of the tests; these, twelve in all, have been given a tied ranking of 18.

Test A	40·80	41·70	36·75	37·55	29·40	25·20	26·75	28·45	26·85	26·35
Test B	1·5	1·5	1·5	2·5	3·5	10	2·5	2·5	2·5	6·5
Test A	21·40	19·65	18·95	22·90	22·80	20·25	24·45	22·70	26·50	—
Test B	>10	>10	>10	>10	>10	>10	>10	>10	>10	—
Test A	22·00	27·50	23·75	30·80	21·00	27·10	22·10	19·25	25·45	24·10
Test B	>10	3·5	1·5	2·5	7	6·5	7·5	9	3·5	>10

Ranking the results according to their order in test A (from highest values to lowest) there is obtained

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	5	1	5	10	5	10	13	5	5	18	13	10	18

16	17	18	19	20	21	22	23	24	25	26	27	28	29
18	18	1	18	18	18	16	18	18	15	18	18	17	18

The lower ranking has 1 pair, 1 triplet, 1 quadruplet, 1 quintuplet and one 12-member multiplet. The maximum possible score is $\frac{29\times 28}{2}-1-3-6-10-66=320$ in such a ranking. The actual score is $+212$, giving $r=\frac{212}{320}=0{\cdot}6625$. The variance of the distribution of the scores obtained with such rankings in the case of no correlation is

$$\frac{29\,.\,28\,.\,63}{18}-1-\frac{11}{3}-\frac{26}{3}-\frac{50}{3}-\frac{638}{3}=2599{\cdot}33.$$

Hence the probability of obtaining a score of 212 or more from an uncorrelated pair of such rankings corresponds to the probability of a normal variate attaining or exceeding $\frac{212}{\sqrt{(2599{\cdot}33)}}=4{\cdot}158$ times its standard deviation.

REFERENCES

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81.

Kendall, M. G. (1942). Partial rank correlation. *Biometrika*, **32**, 277.

Kendall, M. G. (1943). *The Advanced Theory of Statistics*, **1**, chapter 16, Rank Correlation. London: C. Griffin and Co. Ltd.

Kendall, M. G. (1945). The treatment of ties in ranking problems. *Biometrika*, **33**, 239.

THE USE OF RANGE IN PLACE OF STANDARD DEVIATION IN THE t-TEST

BY E. LORD, B.Sc., *Shirley Institute, Didsbury, Manchester*

CONTENTS

(i) INTRODUCTION

The difference between highest and lowest values has always been recognized as a general indication of the variability of quantitative data. It was not, however, until 1925 that attention began to be focused upon the range as a useful statistical tool. In his paper 'On the extreme individuals and the range in samples taken from a normal population', Tippett (1925) obtained an expression for the mean value of the range in repeated random samples, and computed its value in terms of the population standard deviation for samples of size $n = 2$ to $n = 1000$. He also gave numerical approximations to the values of the moments of the range for fairly large samples.

The work was taken up by E. S. Pearson (1926), who determined numerically the exact values of the moments of the random sampling distribution of the range for small samples of size $n \leqslant 6$, and also approximations to their values for samples of medium size. In a subsequent paper E. S. Pearson (1932) tabulated the upper and lower percentage limits for the distribution of the range from frequency curves fitted with the values of the moment coefficients taken from both of the earlier papers cited above.

The next advance was the determination of a general expression for the distribution of the range in samples of n random values from any population by McKay & Pearson (1933). For the normal population, only in the case of $n = 3$ was it found possible to obtain a fairly simple analytical form. (The distribution for $n = 2$ is, of course, well known, taking the form of the positive half of a normal curve.)

Hartley (1942) later determined an expression for the probability integral of the range and, with Pearson (1942), tabulated this for the normal population for samples between $n = 2$ and $n = 20$. This latter paper also contains a table of several percentage limits of the range in samples from a normal population. These limits are derived from the numerical values of the probability integrals and replace the approximate values previously given by Pearson referred to above.

Tippett (1925) and Pearson (1932) have pointed out that although the total range in a sample may be used for the purpose of estimating the population value of the standard

deviation, a more efficient measure may be obtained by dividing the sample into random subgroups of equal size and using the mean range of the subgroups in place of the total range. The efficiency of range estimates of standard deviation is, of course, always less than that of root-mean-square estimates, but the work of Davies & Pearson (1934) and Pearson & Haines (1935) indicates that information is not discarded to any serious extent providing that the number of observations in the subsamples is not greatly in excess of about 10.

As a result of the work outlined above, the range is now of considerable importance in many fields, especially in industrial quality control, where its simplicity has enabled it to be extensively and easily applied to the measurement of fluctuations in the variability of quality of a manufactured article or material.

In the present paper an investigation is made of the use of range estimates of standard deviation in the consideration of the statistical significance of deviations of sample means in normal random sampling theory. This use of range estimates of standard deviation is analogous to the use of root-mean-square estimates in the well-known t-test. Tables are given, at several probability levels, and these may be employed in determining the statistical significance of either the deviation of a sample mean from some fixed or hypothetical population value, or the difference between the means of two samples. These tables may also be used for obtaining rapid estimates of the accuracy of a sample mean from the variation within the sample as measured by the range. The use of range, in place of root-mean-square estimates of standard deviation, in this modified form of the t-test necessarily entails some loss of precision. It will, however, be shown in a future paper that this reduction in accuracy is negligible for all practical purposes. Furthermore, this slight disadvantage of the new test is compensated by its greater simplicity, involving a reduced amount of computing compared with the usual t-test.

The range test is suitable for application to many problems frequently encountered in the treatment of various types of experimental data and in considering the mean character value in small samples in biological experiments. In the industrial field, the range test may be used for detecting changes in mean quality level, especially where the variation is not under strict statistical control or is subject to secular changes, or for determining whether the average level of a batch determined from a sample is in accordance with specification demands. A number of these problems are covered in the examples given at the end of the paper.

(ii) THE t-TEST*

In testing the significance of the deviation of a sample mean $\bar{x}$ from an assumed population value ξ, use is made of the ratio

$$t = \frac{|\bar{x}-\xi|}{s/\sqrt{N}}, \tag{1}$$

where N is the size of the sample and s is the root-mean-square estimate of the population standard deviation determined from the sample. In applying this ratio it is assumed that the N values form a random sample from a normal population of which the mean is ξ, standard deviation σ and the distribution of values of x is given by

$$p(x) = \frac{1}{\sqrt{(2\pi)}\,\sigma} e^{-\frac{(x-\xi)^2}{2\sigma^2}}. \tag{2}$$

* 'Student' (1908), R. A. Fisher (1925).

More generally t may be defined as the ratio

$$t = x/s, \tag{3}$$

where x and s are statistically independent, x being a quantity distributed normally about a mean of zero and s a root-mean-square estimate based on ν degrees of freedom of the standard error of x. Although the use of the tables of the probability integral of t enables the most efficient tests to be made of the various forms of the so-called 'Student's Hypothesis', occasions frequently arise when more rapid tests are desirable, especially if accompanied by only inappreciable loss of accuracy. The calculation of s, depending upon the squaring of numerical quantities, entails a certain amount of labour, especially if tables of squares or a calculating machine are not available. The use of the range, or the mean range determined from random subgroups in a sample, enables very rapid estimates to be made of the population value of the standard deviation σ. In the following section these range estimates are used in place of root-mean-square estimates in a modified form of the t-test.

(iii) THE MODIFIED TEST (u-TEST) BASED ON RANGE

Here we replace the s of 'Student's' ratio by an estimate of σ based on the range. Thus

$$u = u(m,n) = \frac{x}{\bar{w}(m,n)/d_n}, \tag{4}$$

where x is a quantity distributed normally about a mean of zero and $\bar{w}(m,n)$ is the mean value of m ranges w, obtained from m independent samples or subgroups, each containing n observations. The constant d_n, in a commonly used notation,* is the expected value of the range in samples of n, randomly selected from a normal population of unit standard deviation. The ratio $\bar{w}(m,n)/d_n$ is therefore an estimate of the standard error of x obtained from range and, as such, replaces the root-mean-square estimate s used in the ratio $t = x/s$.

Except for a few special cases, it has not been found possible to determine the analytical form of the distribution of u, but several tables of percentage points have been computed for use in testing the various statistical hypotheses normally covered by the t-test. The computation of these tables is considerably simplified by first determining the percentage points of the distribution of the subsidiary quantity

$$q = q(m,n) = \frac{u(m,n)}{d_n} = \frac{x}{\bar{w}(m,n)}, \tag{5}$$

and the multiplying by the corresponding value of d_n to obtain the percentage points of the u distribution.

To simplify the algebraic expressions in what follows, u, $\bar{w}$ and q will be written for $u(m,n)$, $\bar{w}(m,n)$ and $q(m,n)$ where no confusion is involved.

The distribution of both u and q are clearly independent of σ. Hence, without any loss of generality, σ may be taken equal to unity in considering the distributions. The distribution of x will therefore be defined by

$$p(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}. \tag{6}$$

* See, for example, Pearson (1935), pp. 84 and 90.

Furthermore, let the distribution of the range w in a sample of n be $y = p(w)$, that of $\bar{w}$ be $y = p(\bar{w})$, and that of q be $y = p(q)$. Then since x and $\bar{w}$ are defined to be statistically independent, we have the distribution of q given by

$$p(q) = \int p(\bar{w})\, p(x)\, d\bar{w}\, dx, \tag{7}$$

where the integral is to be evaluated over the field of all values of x and $\bar{w}$ subject to the relation (5) and to the conditions:

$$-\infty < x < \infty, \quad 0 \leqslant \bar{w} < \infty. \tag{8}$$

Since x is distributed symmetrically about zero, and $\bar{w}$ is independent of x, the ratio q is also symmetrically distributed about zero. Let q_α be the value of q such that α is the chance that $|q| \geqslant q_\alpha$. The quantity α represents the total area of the two equal tails of the distribution lying outside deviations $\pm q_\alpha$, and we have

$$(1-\alpha) = 2\int_0^{q_\alpha} p(q)\, dq. \tag{9}$$

Alternatively, from (6), (7) and (8), this may be written in the form

$$1-\alpha = \int_0^\infty \left\{ p(\bar{w}) \int_{-\bar{w}q_\alpha}^{+\bar{w}q_\alpha} p(x)\, dx \right\} d\bar{w}$$

$$= 2\int_0^\infty \left\{ p(\bar{w}) \int_0^{\bar{w}q_\alpha} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\, dx \right\} d\bar{w}. \tag{10}$$

Except for a few cases in which the analytical form of the distribution of u has been obtained, equation (10) has been used to compute values of q_α and, hence, the percentage points of u for values of $\alpha = 0{\cdot}10$, $0{\cdot}05$, $0{\cdot}02$, $0{\cdot}01$, $0{\cdot}002$ and $0{\cdot}001$, with values of n from 2 to 20 and values of m from 1 to 10, 15, 20, 30, 60 and 120.

The percentage points of $u = u(m, n)$ are first considered for the case when the estimate of standard deviation is based on the value of a single range of n random values (i.e. for $m = 1$). This treatment is followed by the case of $m = 2$, and finally consideration is given to the general case using estimates of standard deviation determined from the mean of m ranges each from an independent subgroup of n random values.

(iv) Computation of percentage points of the distribution of $u = u(1, n)$, i.e. case with $m = 1$

Throughout this section the estimates of standard deviation are all based upon the value of the range in a single set of n random values of the variate (thus $\bar{w} = w$). In the case of $n = 2$ and $n = 3$ analytical solutions are derived for the distributions from which the percentage points of q and u are calculated. For $n \geqslant 4$, percentage points in the neighbourhood of those desired are determined by quadrature methods, and the required points obtained from these by interpolation.

Special case $n = 2$, $m = 1$

The distribution of ranges (w) in samples of two random values from a normal population with unit standard deviation is the distribution of absolute differences between random pairs of variate values, and this may be easily shown to be

$$p(w)\, dw = \frac{1}{\sqrt{\pi}} e^{-\frac{1}{4}w^2}\, dw \quad (0 \leqslant w < \infty). \tag{11}$$

Since x and w are independent, it follows from (6) and (11) that their joint probability distribution is

$$p(w, x)\,dw\,dx = \frac{1}{\pi\sqrt{2}}\,e^{-\frac{1}{2}(x^2+\frac{1}{2}w^2)}\,dw\,dx. \tag{12}$$

Transforming to new variables, $q = x/w$ and w and noting that the Jacobian of the transformation

$$\frac{\partial(x, w)}{\partial(q, w)} = w,$$

the joint distribution of q and w is given by

$$p(q, w)\,dq\,dw = \frac{1}{\pi\sqrt{2}}\,e^{-\frac{1}{2}w^2(q^2+\frac{1}{2})}\,w\,dq\,dw. \tag{13}$$

To obtain the distribution of q it is necessary to integrate (13) over the whole field of w, from 0 to ∞. This gives

$$p(q)\,dq = \frac{dq}{\pi\sqrt{2}\,(q^2+\frac{1}{2})}. \tag{14}$$

Hence from (9) and (14) above, the percentage points of q are given by

$$\begin{aligned}(1-\alpha) &= \frac{2}{\pi\sqrt{2}}\int_0^{q_\alpha}\frac{dq}{(q^2+\frac{1}{2})}\\ &= \frac{2}{\pi}\tan^{-1}(\sqrt{2}\,q_\alpha),\end{aligned}$$

and hence

$$q_\alpha(1, 2) = \frac{1}{\sqrt{2}}\tan\left\{\frac{\pi}{2}(1-\alpha)\right\} = \frac{1}{\sqrt{2}}\cot\left(\frac{\pi\alpha}{2}\right). \tag{15}$$

The values of q_α determined from (15), for the six values of α under consideration, are multiplied by $d_2 = 2/\sqrt{\pi}$ to give the required percentage points of the distribution of $u = u(1, 2)$.

Special case $n = 3$, $m = 1$

For random samples of size $n = 3$ from a normal distribution with unit standard deviation, the distribution of the range has been found by McKay & Pearson (1933) and takes the form

$$p(w) = \frac{6}{\pi\sqrt{2}}\,e^{-\frac{1}{4}w^2}\int_0^{w/\sqrt{6}} e^{-\frac{1}{2}z^2}\,dz.$$

Again, since w and x are independent, their joint distribution is given by

$$p(w, x)\,dw\,dx = \frac{3}{\pi\sqrt{\pi}}\,e^{-\frac{1}{2}(x^2+\frac{1}{2}w^2)}\,dw\,dx\int_0^{w/\sqrt{6}} e^{-\frac{1}{2}z^2}\,dz. \tag{16}$$

Transforming to new variables $q = x/w$ and w, it follows from (16), since the Jacobian of the transformation is equal to w, that the joint distribution of q and w is given by

$$p(q, w)\,dq\,dw = \frac{3}{\pi\sqrt{\pi}}\,e^{-\frac{1}{2}w^2(q^2+\frac{1}{2})}\,w\,dw\,dq\int_0^{w/\sqrt{6}} e^{-\frac{1}{2}z^2}\,dz. \tag{17}$$

To obtain the distribution of q the expression in (17) has to be integrated over the whole field of w, from 0 to ∞. Thus

$$p(q) = \frac{3}{\pi\sqrt{\pi}}\int_0^\infty e^{-\frac{1}{4}w^2(q^2+\frac{1}{2})}w\,dw\int_0^{w/\sqrt{6}} e^{-\frac{1}{2}z^2}\,dz.$$

Putting $t = z\sqrt{6}$, the above may be written

$$p(q) = \frac{3}{\pi\sqrt{(6\pi)}}\int_0^\infty e^{-\frac{1}{4}w^2(q^2+\frac{1}{2})}\,d(\tfrac{1}{2}w^2)\int_0^{w} e^{-t^2/12}\,dt$$

$$= \frac{-3}{\pi\sqrt{(6\pi)}}\left[\frac{1}{(q^2+\frac{1}{2})}\,e^{-\frac{1}{4}w^2(q^2+\frac{1}{2})}\int_0^{w} e^{-t^2/12}\,dt\right]_{w=0}^{w=\infty}$$

$$+\frac{3}{\pi\sqrt{(6\pi)}\,(q^2+\frac{1}{2})}\int_0^\infty e^{-w^2/12}\,e^{-\frac{1}{4}w^2(q^2+\frac{1}{2})}\,dw. \tag{18}$$

The first expression in (18) is clearly zero and hence

$$p(q) = \frac{3}{\pi\sqrt{(6\pi)}\,(q^2+\frac{1}{2})}\int_0^\infty e^{-\frac{1}{4}w^2(q^2+\frac{1}{2})}\,dw$$

$$= \frac{\sqrt{3}}{2\pi(q^2+\frac{1}{2})\,(q^2+\frac{2}{3})^{\frac{1}{2}}}. \tag{19}$$

As before $$(1-\alpha) = \frac{\sqrt{3}}{\pi}\int_0^{q_\alpha}\frac{dq}{(q^2+\frac{1}{2})\,(q^2+\frac{2}{3})^{\frac{1}{2}}} = \frac{6}{\pi}\tan^{-1}\left\{\frac{q_\alpha}{(2+3q_\alpha^2)^{\frac{1}{2}}}\right\},$$

and therefore $$\frac{q_\alpha}{(2+3q_\alpha^2)^{\frac{1}{2}}} = \tan\left\{\frac{\pi}{6}(1-\alpha)\right\}.$$

If $\tan\left\{\frac{\pi}{6}(1-\alpha)\right\}$ be denoted by τ, then

$$q_\alpha(1,3) = \frac{\sqrt{2}\,\tau}{(1-3\tau^2)^{\frac{1}{2}}}. \tag{20}$$

The six required values of q_α are found by substitution of the corresponding values of α in (20) above, and further multiplication by $d_3 = 3/\sqrt{\pi}$ gives the percentage limits of the distribution of $u = u(1,3)$.

General case $n \geqslant 4$, $m = 1$

For $n \geqslant 4$ no suitable algebraic expression exists for the distribution of the range, but Pearson & Hartley (1942) have tabulated values of the probability integral $\int_0^w p(w)\,dw$ to 4 figures at intervals of 0·05 of w for values of n from 2 to 20. Hartley kindly lent manuscript tables of the integral tabulated to 5 figures at intervals of 0·25 of w for values of n from 2 to 16. Using these five-figure tables for $n = 4, 6, 10, 16$ and the four-figure tables for $n = 20$, the frequency distribution of w was obtained numerically by subtraction of successive values of $\int_0^w p(w)\,dw$ at intervals of 0·25 and then converting these class frequencies into ordinates $y(w)$. The degree of approximation in the formula used implied the vanishing of fifth differences (see K. Pearson, *Tables for Statisticians and Biometricians*, Part II, p. xvii).

Each case was treated in turn, and the six values of the percentage points q_α, corresponding to the six different values of α under consideration, were determined using the relations given in (10) above. Taking a trial value of q_α, the integrals

$$I(w, q_\alpha) = \frac{2}{\sqrt{(2\pi)}} \int_0^{wq_\alpha} e^{-\frac{1}{2}x^2} dx \tag{21}$$

were calculated at intervals of 0·25 over the whole range of w. Quadrature was then applied to the products $y(w)\, I(w, q_\alpha)$ over the range $0 \leqslant w < \infty$, to obtain the value of $(1-\alpha)$ corresponding to the trial value of q_α. This procedure was repeated a number of times to obtain values of $(1-\alpha)$ corresponding to a series of equidistant values of q_α. The required values of q_α corresponding to the six values of α under investigation were then obtained by backward interpolation.

As $n \to \infty$ the ratio w/d_n tends to the population value of the standard deviation. Furthermore, for $n = 2$ and $n = 3$, exact values of q_α had been previously obtained by direct calculation for the six values of α. Thus it was possible to make initial estimates of the required values of q_α, and the process of this 'trial and error' method was not found too laborious.

Table 1. *Framework values of percentage points of* $u = u(1, n)$

n \ α	0·10	0·05	0·02	0·01	0·002	0·001
2	5·0376	10·1381	25·389	50·791	253·97	507·95⁻
3	2·5935⁻	3·8225⁺	6·188	8·819	19·84	28·08
4	2·1793	2·9505⁺	4·213	5·420	9·42	11·75⁺
6	1·9354	2·4755⁺	3·249	3·900	5·71	6·66
10	1·8064	2·2390	2·807	3·244	4·32	4·82
16	1·7496	2·1385⁺	2·628	2·990	3·82	4·19
20 (*a*)	1·7320	2·1083	2·576	2·916	3·69	4·01
20 (*b*)	1·7314	2·1074	2·576	2·916	3·69	4·02
∞	1·6449	1·9600	2·326	2·576	3·09	3·29

The framework values of the percentage points of $u(1, n)$ were obtained by multiplying the values of q_α by the corresponding values of the mean range d_n tabulated by Tippett (1925) and are given in Table 1, together with the exact values for $n = 2$ and $n = 3$ determined above.

As a check on the accuracy of determination of these percentage points, the six values for $n = 20$ were also calculated by a second method. Writing

$$r = 1/q = d_n/x, \tag{22}$$

the method is to determine, for the six values of α, the corresponding values of q_α such that

$$\alpha = \int_{-\infty}^{-1/q_\alpha} p(r)\, dr + \int_{1/q_\alpha}^{\infty} p(r)\, dr, \tag{23}$$

where $y = p(r)$ denotes the frequency distribution of r. Since q is distributed symmetrically about zero, its reciprocal r is also distributed symmetrically about zero, and from (22) and (23) it follows that

$$\alpha = 2 \int_{1/q_\alpha}^{\infty} p(r)\, dr$$

$$= 2 \int_0^\infty p(x)\, dx \int_0^{x/q_\alpha} p(w)\, dw. \tag{24}$$

Ordinates of the normal curve $y(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$ were taken at intervals of 0·25 for x in the range $0 \leqslant x < \infty$ from K. Pearson's *Tables for Statisticians and Biometricians*, Part I. Taking a trial value of $1/q_\alpha$, the integrals $\int_0^{x/q_\alpha} p(w)\,dw$ were calculated from Hartley's four-figure tables for each value of x in the above range. By quadrature applied to the products $p(x)\int_0^{x/q_\alpha} p(w)\,dw$ the integral (24) was evaluated. By taking a series of equidistant values of $1/q_\alpha$ other trial values of α were determined. Backward interpolation was then used to obtain the required values of q_α corresponding to the six values of α under consideration. Finally, the six percentage points of u were determined by multiplying the values of q_α by d_{20} given in Tippett's table. These percentage points are given in the penultimate line (b) of Table 1, and comparison with the corresponding values in the line above, (a), indicates good agreement between the two methods of computation.

Since the percentage points of u for $n = 4$, 6, 10 and 16 have been determined using Hartley's five-figure manuscript tables of the cumulative frequency distribution of w, they should be at least as accurate as the percentage points for $n = 20$ determined from the four-figure tables.

Changes in the percentage points at the different levels of significance run most smoothly if arguments proportional to $1/n$ are used in place of n, and reciprocals of u for the variate. Using a six-point general Lagrangian formula applied to the points corresponding to $n = 3, 4, 6, 10, 16$ and 20, values of percentage points of u were determined for $n = 5, 7, 8, 14$ and 18. (In the case of $n = 20$ the mean values of the percentage points determined by the two methods were used.) The interval was then halved, using a nine-point Lagrangian through points corresponding to $n = 4, 6, 8, 10, 12, 14, 16, 18$ and 20. Finally, the six sets of percentage points were differenced as a check, reduced by either one or two figures and, with the exception of those for $n = 5$, are given in the second columns of Tables 3–8 under $m = 1$.

For $n = 5$, the six percentage points of u were independently determined at a later stage of the investigation by the method used above for $n = 4$, 6, 10, 16 and 20, and it is these values which are given in Tables 3–8. In the table below, the values obtained by interpolation from the framework values are compared with those determined by direct calculation.

Percentage points of $u(1, 5)$

$\alpha =$	0·10	0·05	0·02	0·01	0·002	0·001
By direct calculation	2·019	2·635−	3·56	4·38	6·8	8·2
By interpolation	2·020	2·635+	3·56	4·38	6·8	8·2

In one case only, for $\alpha = 0{\cdot}10$, is there a difference as much as one unit in the last figure, the actual values obtained being 2·0192 by direct computation and 2·0198 by interpolation.

Taking all the various checks into consideration, it appears unlikely that the values of the percentage points of $u(1, n)$ given in the tables at the end of the paper are in error by more than one unit in the last place. The values for $n = 2$ and $n = 3$ are, of course, exact.

(v) COMPUTATION OF PERCENTAGE POINTS OF THE DISTRIBUTION OF $u = u(2, n)$, i.e. CASE WITH $m = 2$

The exact distribution of $u(2,n)$ cannot be determined analytically except in the case of $n = 2$, and hence the various percentage points have necessarily to be evaluated almost wholly by numerical methods.

The probability of the joint occurrence of a pair of ranges w' and w'' from random samples of equal size n from a normal population of unit standard deviation is given by

$$p(w', w'')\,dw'\,dw'' = p(w')\,p(w'')\,dw'\,dw''. \tag{25}$$

If $\overline{w}$ be the mean value of the two ranges, then its distribution is obtained by integrating (25):

$$p(\overline{w})\,d\overline{w} = \int p(w')\,p(w'')\,dw'\,dw'', \tag{26}$$

the integration being taken over the whole field of w' and w'' subject to the conditions

$$\overline{w} = \tfrac{1}{2}(w'+w'') \quad (0 \leqslant w' < \infty,\ 0 \leqslant w'' < \infty). \tag{27}$$

We shall change the variables from w' and w'' to $\overline{w}$ and w', the Jacobian of the transformation being equal to 2. With these new variables, and noting from (27) that w' varies from 0 to $2\overline{w}$, equation (26) gives

$$p(\overline{w}) = 2\int_0^{2\overline{w}} p(w')\,p(2\overline{w}-w')\,dw'. \tag{28}$$

Special case $n = 2$, $m = 2$

In equation (11) above is given the distribution of the range in samples of two random values from a normal population with unit standard deviation. Substituting this in (28) above, the distribution of means of two independent values of w is therefore given by

$$\begin{aligned} p(\overline{w}) &= 2\int_0^{2\overline{w}} \frac{1}{\sqrt{\pi}}\,e^{-\frac{1}{4}w'^2}\,\frac{1}{\sqrt{\pi}}\,e^{-\frac{1}{4}(2\overline{w}-w')^2}\,dw' \\ &= \frac{2}{\pi}\,e^{-\frac{1}{2}\overline{w}^2}\int_0^{2\overline{w}} e^{-\frac{1}{2}(w'-\overline{w})^2}\,dw' \\ &= \frac{8}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}\overline{w}^2}\int_0^{2\overline{w}} \frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}z^2}\,dz. \end{aligned} \tag{29}$$

Using the notation

$$I(\overline{w}) = \int_0^{\overline{w}} \frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}z^2}\,dz,$$

the expression (29) for the distribution of the mean range may be written in the form

$$p(\overline{w}) = \frac{8}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}\overline{w}^2}\,I(\overline{w}). \tag{30}$$

We may now proceed to determine the distribution of the ratio $q = x/\overline{w}$. Since they are independent, the joint distribution x and $\overline{w}$ is, from (6) and (30), given by

$$p(x,\overline{w})\,dx\,d\overline{w} = \frac{4}{\pi}\,e^{-\frac{1}{2}x^2}\,e^{-\frac{1}{2}\overline{w}^2}\,I(\overline{w})\,dx\,d\overline{w}. \tag{31}$$

Transforming to variables q and $\bar{w}$, noting that the Jacobian of the transformation is equal to $\bar{w}$, and integrating, we obtain

$$p(q) = \frac{4}{\pi}\int_0^\infty e^{-\frac{1}{4}\bar{w}^2(q^2+1)}\, I(\bar{w})\,\bar{w}\,d\bar{w}. \tag{32}$$

Now since
$$\frac{dI(\bar{w})}{d\bar{w}} = \frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{4}\bar{w}^2},$$

we make use of the identity

$$\frac{d}{d\bar{w}}\{e^{-\frac{1}{2}\bar{w}^2(q^2+1)}\,I(\bar{w})\} = -\bar{w}(q^2+1)\,e^{-\frac{1}{2}\bar{w}^2(q^2+1)}\,I(\bar{w}) + \frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{4}\bar{w}^2}\,e^{-\frac{1}{2}\bar{w}^2(q^2+1)}$$

to evaluate the integral in (32) and obtain

$$p(q) = \frac{4}{\pi\sqrt{(2\pi)}\,(q^2+1)}\int_0^\infty e^{-\frac{1}{4}\bar{w}^2}\,e^{-\frac{1}{2}\bar{w}^2(q^2+1)}\,d\bar{w}$$

$$= \frac{2}{\pi(q^2+1)\sqrt{(q^2+2)}}. \tag{33}$$

The ratio q is distributed symmetrically about zero and from (9) and (33) we obtain

$$1-\alpha = \frac{4}{\pi}\int_0^{q_\alpha}\frac{dq}{(q^2+1)\sqrt{(q^2+2)}},$$

and the percentage points are therefore given by

$$q_\alpha = q_\alpha(2,2) = \frac{\tau\sqrt{2}}{1-\tau}, \tag{34}$$

where
$$\tau = \tan\left\{\frac{\pi}{4}(1-\alpha)\right\}.$$

Substitution of $\alpha = 0{\cdot}10$, $0{\cdot}05$, etc., in (34) gives the required values of q_α, and further multiplication by $d_2 = 2/\sqrt{\pi}$ gives the required corresponding percentage points of the distribution of $u(2,2)$.

General case $n \geqslant 3$, $m = 2$

For $n = 4$, 6, 10, 16 and 20, the ordinates $y(w)$, of the distribution of the range have been previously evaluated above at intervals of 0·25 for w, and these are used in place of the unknown $p(w)$. Taking a particular value of $\bar{w}$, quadrature was then applied to the products $y(w)\,y(2\bar{w}-w)$ to obtain a numerical estimate of $p(\bar{w})$ from equation (28). This process was repeated at intervals of 0·25 for $\bar{w}$ through as much of the range $0 \leqslant \bar{w} < \infty$ as was necessary to obtain the required degree of accuracy. For $n = 3$, determination of the ordinates of the distribution of $\bar{w}$ was also based on quadrature, but, in this case, exact figures for the ordinates of the distribution of the range were obtained from its equation found by McKay & Pearson (1933). Because of the rapid rise of the distribution near to the origin, estimates of the lower values of $p(\bar{w})$ in this case were determined using an interval of $\frac{1}{16}$ for $\bar{w}$ in order to obtain the requisite accuracy. For higher values of $\bar{w}$ the interval was progressively decreased to 0·25 used over the tail portion of the curve.

Treating in turn each curve of the distribution of $\overline{w}$ (for $n = 3, 4, 6, 10, 16$ and 20), the values of the percentage points of the distribution of the ratio

$$q = \frac{u}{d_n} = \frac{x}{\overline{w}}$$

were computed by a method similar to that used in evaluating the percentage points of q for the case $m = 1$. Taking trial values of q_α, the integrals

$$I(\overline{w}, q_\alpha) = \frac{2}{\sqrt{(2\pi)}} \int_0^{\overline{w}q_\alpha} e^{-\frac{1}{2}x^2}\,dx$$

were calculated at intervals of 0·25 for $\overline{w}$. Quadrature was then applied to the products $y(\overline{w})\, I(\overline{w}, q_\alpha)$ over the range $0 \leqslant \overline{w} < \infty$ to obtain corresponding values of $(1-\alpha)$, the sum of the two tails of the distribution beyond deviations $\pm q_\alpha$. Repeating this procedure, a series of values of $(1-\alpha)$ were obtained, corresponding to a set of equidistant values of q_α. Backward interpolation was then used to obtain the six values of q_α corresponding to the six values of α under consideration. Finally, the required percentage points of u were obtained

Table 2. *Framework values of percentage points of* $u = u(2, n)$

n \ α	0·10	0·05	0·02	0·01	0·002	0·001
2	2·6203	3·8671	6·266	8·932	20·10	28·47
3	2·0201	2·6365+	3·555+	4·340	6·78	8·66
4	1·8760	2·3672	3·047	3·600	5·07	5·80
6	1·7791	2·1920	2·727	3·133	4·11	4·52
10	1·7225+	2·0926	2·551	2·884	3·64	3·96
16	1·6961	2·0470	2·472	2·775+	3·44	3·71
20	1·6879	2·0329	2·449	2·742	3·38	3·64

by multiplication by the appropriate value of d_n, and are given in Table 2, together with the exact values for $n = 2$ determined from (34).

As before, Lagrangian formulae were used to interpolate the intermediate values of the percentage points of u, again taking arguments proportional to $1/n$ and reciprocals of the percentage points for the variate. As a check, the values were inspected by determining differences up to the third order and then reduced by one place of decimals. With the exception of the values for $n = 5$, the reduced values are given in Tables 3–8.

For $n = 5$ the six percentage points of u were independently determined at a later stage of the investigation by the same methods used for the framework values. These directly computed values are given in the tables mentioned above, and are also reproduced below for comparison with those obtained by interpolation from the framework values.

Percentage points of $u = u(2, 5)$

$\alpha =$	0·10	0·05	0·02	0·01	0·002	0·001
By direct calculation	1·814	2·254	2·84	3·29	4·4	5·0
By interpolation	1·814	2·254	2·84	3·29	4·4	4·9

In the case of $\alpha = 0{\cdot}001$, direct calculation gives a value of 4·97 compared with 4·92 obtained by interpolation. For other percentage levels the agreement is exact to the number of figures quoted.

(vi) Computation of percentage points of the distribution of $u = u(m, n)$ for $m > 2$

The variance of the mean range in m subgroups of equal size n steadily decreases as m increases, and the ratio $\bar{w}(m,n)/d_n$ gives closer estimates of the population value of the standard deviation of the variate. Hence, following the usual methods of large-sample theory, the limiting values of the percentage points of u, for indefinitely large m, may be determined from integral tables of the normal curves. For a given value of α, the limiting values of the percentage points are, of course, equal for all values of n, and are also equal to the corresponding limiting values of the percentage points of Fisher's t-distribution for an indefinitely large number of degrees of freedom.

In general it was found for a particular value of α and of n that a three-point Lagrangian curve with $1/m$ as argument and reciprocals of the percentage points of u as variate (passing through points corresponding to $m = 1$, 2 and ∞) may be used for interpolation of the required percentage points corresponding to values of m intermediate between 2 and ∞. Only in the case of $n = 2$ and $n = 3$ was the required accuracy not attained by this procedure and further investigation found necessary. Details of the methods used are given below.

In the case of $n = 2$, the percentage points of the distribution of $u(m, 2)$ were also determined for $m = 4$ and $m = 8$ as follows. First, considering $m = 4$, it was necessary to obtain numerical estimates of the ordinates of the distribution of the means of four ranges, each range from a random sample of two values from a normal population with unit standard deviation. Following the method used for $m = 2$ and leading to equation (28), it is easy to show that the distribution of $\bar{w}(2m, n)$, the mean of $2m$ independent ranges each from a random subsample of n values, is given in terms of the distribution of the mean, $\bar{w}(m, n)$, of m such ranges by

$$p(\bar{w}(2m,n)) = 2\int_0^{2\bar{w}(2m,\,n)} p(\bar{w}(m,n))\,p(2\bar{w}(2m,n)-\bar{w}(m,n))\,d\bar{w}(m,n). \tag{35}$$

Using numerical values for $p(\bar{w})$ $(m = 2,\ n = 2)$ given by equation (29) above, estimates of the ordinates of the distribution of $p(\bar{w})$ for $m = 4$, $n = 2$ at intervals of 0·25 were found by quadrature methods similar to those described in previous sections by using the above expression. A repetition of this process, using these last computed values, yielded numerical estimates of the distribution of the means of eight ranges, i.e. $m = 8$. Again applying quadrature to the two distributions, values of $(1-\alpha)$ were determined for a series of equidistant values of q_α $(m = 4$ and $m = 8)$. The required values of q_α, corresponding to the six values of α between 0·10 and 0·001 under consideration, were then obtained by backward interpolation, and hence the percentage points of $u(4, 2)$ and $u(8, 2)$.

The sets of six percentage points of $u(m, 2)$ were determined for each required value of m by Lagrangian interpolation, reciprocals of m being used as argument and reciprocals of the corresponding percentage points as variate, the curve passing through the points corresponding to $m = 1$, 2, 4, 8 and ∞. The interpolated values of percentage points obtained by this method, and the directly computed values for $m = 4$ and $m = 8$, are given in Tables 3–8

for a series of values of m suitable for practical use. As a check upon the method, the computation was repeated, this time using a four-point Lagrangian passing through points corresponding to $m = 1, 2, 4$ and ∞. Most of these interpolated four-point values of the percentage points agree exactly with the five-point values previously obtained. In cases where differences arise, none exceed 1 unit in the last figure. The five-point Lagrangian method of interpolation therefore certainly appears to be quite adequate for furnishing the required degree of accuracy.

Numerical estimates of the distribution of the means of pairs of ranges from subsamples of size $n = 3$ and $n = 4$ have already been obtained above. Using these in turn in equation (35), estimates of the ordinates of the distributions of the means of four ranges were determined at intervals of 0·25 by quadrature methods. The sets of percentage points of $u(4, 3)$ and $u(4, 4)$ were then computed by the previous method of trial values and subsequent backward interpolation.

For $n = 3$ and $n = 4$, the percentage points of the distribution of u, for given values of m, are lower and nearer to their limiting values than the corresponding points for $n = 2$. Furthermore, the changes in the values of the percentage points for small values of m are also less abrupt. In view of the agreement between the four-point and five-point Lagrangian interpolated values of the percentage points for $n = 2$, a four-point Lagrangian through points corresponding to $m = 1, 2, 4$ and ∞ may certainly be relied upon to give adequate accuracy for the interpolation of percentage points corresponding to intermediate values of m in the case of $n = 3$ and $n = 4$. These values, together with the computed values for $m = 4$, are given in Tables 3–8.

For the remaining values of n, from 5 to 20, the interpolated percentage points given in Tables 3–8 have been obtained by means of a three-point Lagrangian curve, using values of the percentage points corresponding to values of $m = 1, 2$ and ∞. As in the previous cases of interpolation, reciprocals of m and the variate were used in order to obtain small changes in successive differences. To show that this method is adequate, the six sets of percentage points for $n = 4$ were also interpolated using a three-point Lagrangian. In every case except one, these values agreed exactly with the four-point Lagrangian interpolated values previously found and given in Tables 3–8. In the case of the sole exception, the difference between the two interpolated values was only one unit in the last figure. For the less rapidly changing values of the percentage points of u for $n \geqslant 5$, the three-point method of interpolation therefore provides sufficient accuracy for the present purpose.

Taking all checks into consideration it appears that the tabulated values of the percentage points of the distribution of the function $u = u(m, n)$ may be relied upon to the accuracy given: occasionally the values may be one unit in error in the last figure.

In Tables 3 and 4, the 10 % and 5 % points of u were computed to 3 decimal places, but lack of space has necessitated these being curtailed for publication. For the same reason the values of the percentage points of u for the odd values of $n = 11, 13 \ldots 19$ have been omitted. In practical applications of the test, it is not considered that this reduction will cause any undue inconvenience. Fuller tables have, however, been retained for forthcoming work on the power of the u-test and are available for consultation if required.

(vii) APPROXIMATE VALUES OF THE PERCENTAGE POINTS OF u

If there are m subgroups each of n values, and if the estimate of standard deviation is determined as the root-mean-square of the deviations of variate values from the respective

means of the subgroups, then the number of degrees of freedom is $\nu = m(n-1)$. Unlike the usual t-test, when the estimate of standard deviation is determined from the mean range in m subgroups of equal size n, the percentage points of the modified t-distribution investigated above depend upon the relation between m and n. Reference to Tables 3–8 indicates that, for a constant number of degrees of freedom $\nu = m(n-1)$, the values of the percentage points on a given probability level vary slightly as m and n vary. For example, taking $\alpha = 0{\cdot}05$ and $\nu = 8$, we have the following percentage points: 2·272 for $m = 1$ and $n = 9$, 2·254 for $m = 2$ and $n = 5$, 2·250 for $m = 4$ and $n = 3$, and 2·264 for $m = 8$ and $n = 2$. In general, however, the range in the values of the percentage points of u for a given value of ν is small, and this permits the construction of a table giving approximate values of the six sets of percentage points corresponding to different numbers of degrees of freedom.

Approximate values of percentage points of u

Degrees of freedom $\nu = m(n-1)$	Values of α					
	0·10	0·05	0·02	0·01	0·002	0·001
1	5·0	10·1	25·4	50·8	254·0	507·9
2	2·6	3·8	6·2	8·9	19·9	28·3
3	2·2	3·0	4·2	5·5	9·4	11·8
4	2·0	2·6	3·6	4·4	6·8	8·3
5	1·9	2·5	3·3	3·9	5·7	6·7
6	1·9	2·4	3·1	3·6	5·1	5·9
7	1·8	2·3	3·0	3·5	4·7	5·4
8	1·8	2·3	2·9	3·3	4·5	5·0
9	1·8	2·2	2·8	3·2	4·3	4·8
10	1·8	2·2	2·7	3·1	4·2	4·6
11	1·8	2·2	2·7	3·1	4·1	4·5
12	1·8	2·2	2·7	3·1	4·0	4·3
13	1·8	2·2	2·7	3·1	3·9	4·3
14	1·7	2·1	2·6	3·0	3·8	4·2
15	1·7	2·1	2·6	2·9	3·7	4·1
16	1·7	2·1	2·6	2·9	3·7	4·1
17	1·7	2·1	2·6	2·9	3·7	4·1
18	1·7	2·1	2·6	2·9	3·6	4·0
19	1·7	2·1	2·6	2·9	3·6	4·0
20	1·7	2·1	2·5	2·8	3·6	3·9
30	1·7	2·0	2·4	2·7	3·4	3·6
60	1·7	2·0	2·4	2·7	3·2	3·5
120	1·7	2·0	2·4	2·6	3·2	3·4
∞	1·64	1·96	2·33	2·58	3·09	3·29

For a particular pair of values of m and n, the values of the percentage points for $\nu = m(n-1)$ degrees of freedom given in the table above are generally not in error by more than one unit in the last place of figures. This degree of accuracy is frequently sufficient for many practical applications of the distribution of u. To settle the significance of cases giving values of u close to the above approximate values, reference should be made to the accurate values given in Tables 3–8.

(viii) Applications of the u-test

The difference between the mean of a sample of n random values of a normally distributed variate and the population value is shown in the Appendix to be independent of the total

range in the sample, and also independent of the mean range determined from random subgroups of values. The modified t-test based on range estimates of standard deviation may therefore be used in various statistical tests of significance involving deviations of sample means. The application of this range test to sampling problems is analogous to that of the well-known t-test, and no detailed description is therefore required. The most frequent use of the new test will be found in the treatment of experimental data of various types, and also in the examination of test results recorded for the purpose of control of the quality of industrial products. In this latter type of work, cases frequently arise when it is desirable to apply a rapid test for determining the significance of a difference between the mean of a sample and some preassigned value, frequently some desired control level, or the significance of the difference between two sample means. Furthermore, for routine purposes, it is often desirable that the test should not only be rapid but also of a simple nature, thus enabling it to be used by workers with little mathematical or even arithmetical aptitude. The new range test has the advantages of greater simplicity and greatly reduced amount of computing compared with the standard t-test. The use of range estimates of standard deviation, in place of root-mean-square estimates, necessarily entails some loss of precision, but in a future paper it will be shown that this reduction in accuracy is small and certainly negligible for most practical purposes.

The most frequent applications of the range test are considered below and are followed by several numerical examples in which, for purposes of comparison, the parallel treatment by the t-test is also given. As in the t-test, the application of the range test involves the assumption of normality of variate distribution and randomness of sampling. Furthermore, where the standard deviation is estimated from the mean range of several subgroups of values, care should be taken to ensure that the arrangement of these values is also random. This latter condition is usually fulfilled by considering the values in the order in which they were originally recorded. In a few cases, however, the order of recording may not be random; the particular circumstances of a test may be such that the order of the observations may be wholly or partly dependent upon their magnitude. In such cases a set of values can be divided into random subgroups by the use of tables of random sampling numbers or by other means.

(a) Difference between sample mean and population mean

Suppose we have some preassigned value ξ, and wish to test whether the mean $\bar{x}$ of a sample of N values may be considered as a reasonable estimate of ξ, or whether the difference between $\bar{x}$ and ξ is real in the statistical sense. The usual assumption, the so-called 'Student's Hypothesis', is made that $\bar{x}$ is the mean of a random sample from a normal population of which the mean is ξ and standard deviation is σ. The differences $(\bar{x}-\xi)$ will be distributed about a mean of zero with a standard error equal to $\sigma/\sqrt{N}$. If the sample be divided into m random subgroups of equal size n, $N = mn$, and $\bar{w}$ is the mean of the m ranges of the subgroups, then the sample estimate of the standard error of the mean is $\bar{w}/(d_n\sqrt{N})$. The ratio of the difference between the means to the estimate of its standard error is

$$u = \frac{|\bar{x}-\xi|\, d_n\sqrt{N}}{\bar{w}}. \tag{36}$$

If the computed value of u exceeds the corresponding percentage point in one of Tables 3–8, then the difference is considered unlikely to have arisen through random sampling on

that particular probability level α. As in the case of the t-test, when considering the asymmetrical case of 'Student's Hypothesis', the values of α at the headings of the tables should be halved.

For fairly small values of N, the estimate of the standard error of the mean may be determined, not from the mean range in subgroups, but from the total range between the maximum and minimum values in the sample. In the notation used above, this corresponds to $m = 1$ and $n = N$. The test of the significance of the difference may be made as above and the computed value of u compared with the percentage points in Tables 3–8. In these cases, however, the computation may be curtailed by using the ratio

$$\frac{\delta}{w} = \frac{u(1, n)}{d_n \sqrt{n}}, \tag{37}$$

where $|\bar{x} - \xi| = \delta$, and w is the range in the undivided sample. Table 9 gives values of the ratio δ/w for various levels of significance corresponding to the sum of the two tails of the distribution. For a chosen level of significance the difference δ is considered too large to have arisen through random sampling errors if the value of δ/w exceeds the corresponding tabulated value. Table 9 will also be found useful for giving a rapid estimate of the accuracy of the mean based on a small number of observations.

(*b*) *Difference between two sample means*

Suppose the first sample of size N_1 be divided into m_1 random subgroups of size n, and the second sample of size N_2 be divided into m_2 random subgroups also of size n, i.e.

$$n = N_1/m_1 = N_2/m_2.$$

The hypothesis is made that each sample can be considered as a random selection from the same normal population. Let the numerical value of the difference between the two sample means be $|\bar{x}_1 - \bar{x}_2|$, and the mean of the $(m_1 + m_2)$ ranges of n values be $\bar{w} = \bar{w}(m_1 + m_2, n)$, giving an estimate $\bar{w}/d_n$ for the standard deviation of the variate. The ratio of the difference between the two sample means to the range estimate of the standard error of the difference is

$$u = \frac{|\bar{x}_1 - \bar{x}_2| \, d_n}{\bar{w}\sqrt{(1/N_1 + 1/N_2)}}. \tag{38}$$

The significance of the difference between the means in any particular case can be determined by noting whether the computed value of u exceeds the corresponding percentage point for a chosen value of α by reference to Tables 3–8, using the column headed $m = m_1 + m_2$.

When the samples are small and of equal size, say n, the variate standard deviation can be estimated from the two total ranges in the samples. If w' and w'' are the two ranges, with a mean value $\bar{w} = \frac{1}{2}(w' + w'')$, then

$$u = \frac{|\bar{x}_1 - \bar{x}_2| \, d_n \sqrt{(\frac{1}{2}n)}}{\bar{w}} \tag{39}$$

may be used as above for testing the significance of the difference between the two means. A more rapid test may, however, be made by simply determining the value of the ratio of the difference between sample means to the average of the two sample ranges

$$\frac{|\bar{x}_1 - \bar{x}_2|}{\frac{1}{2}(w' + w'')} = \frac{u(2, n)}{d_n \sqrt{(\frac{1}{2}n)}}. \tag{40}$$

In Table 10 are given values of the above ratio lying on six different probability levels. For

a given level of significance α, values of the ratio smaller than those tabulated may be considered to have arisen through random sampling errors; greater values indicate that a given difference is unlikely to have arisen through chance and therefore point to a real difference.

In the computation of u it is necessary to use values of d_n, the mean range in samples from a normal population of unit standard deviation. A selection of the values determined by Tippett (1925) is reproduced in Table 11 to avoid the necessity of frequent reference to his original paper, and is accompanied by the corresponding values of $\sqrt{n}$ and $d_n\sqrt{n}$.

(c) *Confidence intervals*

As with 'Student's' test, the tables of percentage points may be used to estimate with a given measure of confidence, the interval within which it can be stated that ξ or $\xi_1 - \xi_2$ lies.

Examples

Example 1. The following data have been previously used as an example by 'Student' (1908). Ten patients were treated with the optical isomers of hyoscyamine hydrobromide and the additional hours of sleep were noted.

Additional hours sleep gained by use of hyoscyamine hydrobromide

Patient	*Dextro*-(D)	*Laevo*-(L)	Difference ($D-L$)
1	+0·7	+1·9	+1·2
2	−1·6	+0·8	+2·4
3	−0·2	+1·1	+1·3
4	−1·2	+0·1	+1·3
5	−0·1	−0·1	0·0
6	+3·4	+4·4	+1·0
7	+3·7	+5·5	+1·8
8	+0·8	+1·6	+0·8
9	0·0	+4·6	+4·6
10	+2·0	+3·4	+1·4
Means	+0·75	+2·33	+1·58

The last column may be used for the controlled comparison of the two drugs, since their effects were measured on the same ten patients. The *laevo* form has given a greater figure for the additional hours sleep than the *dextro* form. Whether the former may be considered as the better soporific is examined by both the standard deviation and range tests.

(a) The sum of squares of deviations of the differences about their mean value is 13·616, associated with 9 degrees of freedom. The estimate of the standard error is therefore 0·3890, and the value of t works out to be $1{\cdot}58/0{\cdot}3890 = 4{\cdot}06$. For 9 degrees of freedom a value of $t = 3{\cdot}250$ lies on the 1 % level of significance. Assuming normal random sampling, a value of t equal or greater than 4·06 will occur much less frequently than once in a hundred times. This leads to the conclusion that the *laevo* form is better for producing sleep than the *dextro* form.

(b) For examination by the range, the value of $u = u(1, 10)$ may be computed, but in this case it is simpler to use the shortened method of equation (37). The ratio of the mean difference to the range in the ten individual differences is $\delta/w = 1{\cdot}58/4{\cdot}6 = 0{\cdot}34$. Reference to Table 9 shows that this value is slightly in excess of the tabulated value 0·333 on the 1 % level of significance, leading to the same conclusion as that drawn from the t-test.

The greater significance suggested by the t-test seems to be largely due to the exceptional difference $D-L$ for Patient No. 9, viz 4·6, which affects s more seriously than w.

Example 2. In the calibration of a viscometer it is necessary to time the interval required for the level of an aqueous solution of glycerol to fall between two fixed marks. For satisfactory calibration it is considered desirable that the mean time of flow should be accurate to $\pm\frac{1}{2}$ sec., risking a greater error not more frequently than 1 in 20 times. Five independent determinations of the time interval (in seconds) for one viscometer were 103·5, 104·1, 102·7, 103·2 and 102·6. While this number of observations is clearly too small for a final assessment of accuracy, it is often useful to get an interim answer to guide further action.

(*a*) The sum of squares of the deviations of the five observations about their mean is 1·508, associated with 4 degrees of freedom, giving an estimate of the standard error of the mean equal to 0·275. Reference to tables shows that a value of t equal to 2·776 lies on the 5 % level of significance. Hence in 19 times out of 20 it would be expected that a sample mean will not diverge from the true mean value by more than $\pm 2{\cdot}776\times 0{\cdot}275 = \pm 0{\cdot}76$ sec. This error exceeds the assigned limits of $\pm\frac{1}{2}$ sec. and therefore points to the necessity of further tests to fulfil the required conditions.

(*b*) Instead of computing an estimate of the standard error of the mean from the range ($w = 104{\cdot}1 - 102{\cdot}6 = 1{\cdot}5$) in the five determinations, we note from Table 9 that a value of $\delta/w = 0{\cdot}507$ lies on the 5 % level of significance. Hence in 19 times out of 20 the sample mean will differ from its true value by an amount up to a deviation of

$$\pm\delta = \pm\, 0{\cdot}507\times 1{\cdot}5 = \pm 0{\cdot}76,$$

a result in agreement with that yielded by the t-test.

Example 3. In the processing of raw cotton, modifications were made in the design of one of the machines with the object of improving the efficiency of cleaning. Tests were made on a series of 24 different mixings for the purpose of determining whether yarn strength was adversely affected by the mechanical alterations. The results of the 24 pairs of comparisons are given below (the strength being expressed as a count × strength product), together with the differences between them expressed as percentages of the corresponding strengths under standard conditions.

Yarn strengths under standard and modified conditions

Strength		Percentage difference	Strength		Percentage difference
Standard S	Modified M	$100(M-S)/S$	Standard S	Modified M	$100(M-S)/S$
1805	1763	−2·3	1931	1898	−1·7
1870	1901	+1·7	1508	1520	+0·8
2000	2026	+1·3	2111	2119	+0·4
1823	1904	+4·4	1496	1481	−1·0
1603	1619	+1·0	1672	1723	+3·1
1889	1830	−3·1	1947	1759	−9·7
2058	2019	−1·9	1960	1934	−1·3
1806	1850	+2·4	1624	1594	−1·8
1056	1112	+5·3	2162	2170	+0·4
1857	1782	−4·0	1915	1967	+2·7
1801	1720	−4·5	1738	1810	+4·1
2094	2144	+2·4	1609	1613	+0·2

The mean value for the percentage difference in strength is $-0{\cdot}46$. Whether this is an indication that the mechanical modifications have resulted in the production of weaker yarns is examined by means of the standard deviation and range tests.

(*a*) The sum of squares of the deviations of the percentage differences about their mean value is 256·48, based on 23 degrees of freedom. The estimate of the standard deviation of the percentage differences is 3·34 and the standard error of their mean value is 0·68, giving a value of $t = 0{\cdot}46/0{\cdot}68 = 0{\cdot}59$. This is much below the value of 2·069 on the 5 % level of significance and leads to the conclusion that there are no grounds for suspecting that the mechanical alterations have led to the production of weaker yarns.

(*b*) The number of observations place this case outside the range of Table 9, and it is therefore necessary to use the modified t-function. The data are arranged in random order of their occurrence, and split into four groups of six, The ranges in the sets of six differences are 7·5, 9·8, 12·8 and 5·9 with a mean value $\overline{w}(4, 6) = 9{\cdot}0$. The estimate of the variate standard deviation is $\overline{w}(4, 6)/d_6 = 3{\cdot}55$, giving 0·72 for the standard error of the mean percentage difference, and $u = 0{\cdot}46/0{\cdot}72 = 0{\cdot}64$. The 5 % level of significance is, from Table 4, equal to 2·07, much greater than the value computed from the data and therefore indicates the same conclusion as above.

Example 4. Independent determinations of percentage trash content were made in triplicate on two samples of raw cotton and the following results obtained:

Percentage trash content of raw cotton

Sample A	Sample B
1·13	0·76
1·31	0·64
1·25	1·01
Means 1·23	0·80

The point to be decided is whether sample B may be said to be cleaner than sample A, or whether the difference between the two average percentage trash contents may be accounted for by random experimental variation. Since, in this case, the comparisons are not paired, the standard error of the difference between the mean values of the two samples has necessarily to be estimated from the variation within each of the two sets of results. As before, normal variation in sampling and in testing errors is assumed.

(*a*) The sum of squares of deviations of each set of values from their mean is 0·01680 for A and 0·07167 for B, each associated with 2 degrees of freedom. The best estimate of the error standard deviation is therefore 0·149, giving 0·122 for the estimate of the standard error of the difference between the two means. The value of t is equal to $(1{\cdot}23-0{\cdot}80)/0{\cdot}122 = 3{\cdot}5$ which exceeds the value 2·776 obtained from tables for 4 degrees of freedom and $\alpha = 0{\cdot}05$. On this level of significance the result is taken to indicate a real difference in the cleanliness of the two cottons.

(*b*) The difference between the two sample means is 0·43 and the mean of the two ranges is 0·275. Hence, using the ratio of equation (40), $|\bar{x}_1-\bar{x}_2|/\frac{1}{2}(w'+w'') = 1{\cdot}6$ which, from Table 10, is seen to exceed the value of 1·272 lying on the 5 % level and therefore is taken to indicate a significant difference in the mean values.

Example 5. The following strength test results were obtained on two batches of cotton yarn (measurements recorded to the nearest ½ lb.) and are noted downwards in order of random occurrence:

Sample *A*			Sample *B*	
30·5	31·0	29·5	27·0	28·5
28·0	31·5	27·5	28·5	25·0
29·5	30·0	28·0	26·5	28·0
28·0	27·5	26·0	27·0	27·5
28·5	29·5	28·5	27·0	27·5
29·5	27·5	26·5	28·5	28·5
27·5	28·0	27·0	28·0	28·0
28·0	32·5	30·0	28·0	26·0
29·5	28·5	28·5	25·0	26·5
30·5	29·0	31·0	29·0	28·0

The mean of sample *A* is 28·90 lb. and 27·40 lb. for sample *B*, and the question arises as to whether sample *B* is actually weaker than *A*.

(*a*) The sums of squares of the deviations about their respective mean values are 68·2 for *A* and 24·8 for *B*, associated with 29 and 19 degrees of freedom. The estimate of the error standard deviation is therefore 1·392, giving 0·402 for the standard error of the difference between the two means and a value of t equal to $(28{\cdot}9-27{\cdot}4)/0{\cdot}402 = 3{\cdot}7$. For 48 degrees of freedom a value of $t = 2{\cdot}68$ lies on the 1 % level of significance. The greater value of 3·7 yielded by the data above indicates, therefore, that the difference in strength of the two yarns may be accepted as statistically significant.

(*b*) The estimate of the error standard deviation is obtained from the ranges within groups of ten values, three groups for sample *A* and two for sample *B*. The values of these five ranges are 3·0, 5·0, 5·0, 4·0 and 3·5 with a mean value of 4·1 and a corresponding estimate of error standard deviation equal to $\bar{w}(5,10)/d_{10} = 1{\cdot}33$. The estimate of the standard error of the difference between the two means is 0·384 and the value of u is $(28{\cdot}9-27{\cdot}4)/0{\cdot}384 = 3{\cdot}9$. For five ranges of ten, the value of u on the 1 % level of significance is, from Table 6, equal to 2·69 (cf. 2·68 for t with 48 degrees of freedom). The value of 3·9 obtained from the data is greater than this value of 2·69 and this again leads to the conclusion that the difference in mean strengths of the two yarns is 'statistically significant'.

Note added in proof. Since the present paper went to press, a note by Daly (1946) has been published, in which it is suggested that the range may be used in place of the root-mean-square estimate of variance in a test analogous to the t-test. The case where the estimate of standard deviation from a single range is discussed and values of the ratio (deviation)/(range) on the 10 % level of significance are given to two significant figures for a number of low values of n. These agree with the corresponding values given in Table 9, for $\alpha = 0{\cdot}10$, of the present paper. [Mr Lord's paper was first submitted for publication in August 1945. Ed.]

APPENDIX

On the independence of mean and some linear estimates of standard deviation in random samples from a normal population

In the above practical applications of the u distribution to normal random sampling problems, it has been implicitly assumed that range estimates of standard deviation, like root-mean-square estimates, are independent of the mean of the sample from which they have been determined. The validity of this assumption is established below, where it is shown as a particular case of a more general theorem.

Consider a set of n random values of a variable from a normal population of distribution

$$p(x)\,dx = \frac{1}{\sqrt{(2\pi)}\,\sigma}\exp\left[-\frac{1}{2}\frac{(x-\xi)^2}{\sigma^2}\right]dx. \tag{1}$$

Of such a set let x_p and x_q denote the pth and qth values ($p<q$) in ascending order of magnitude, and denote the remaining $(n-2)$ values such that

$$\left.\begin{aligned} -\infty < x_r \leqslant x_p, &\quad r = 1, 2, \ldots, (p-1), \\ x_p \leqslant x_r \leqslant x_q, &\quad r = (p+1), (p+2), \ldots, (q-1), \\ x_q \leqslant x_r < \infty, &\quad r = (q+1), (q+2), \ldots, n. \end{aligned}\right\} \tag{2}$$

Now a set of any n values may be arranged in $n!$ ways and in random samples all arrangements of the same n values are of equal probability. The group of $(p-1)$ values all less than x_p are not ranked in any particular order, and there are hence $(p-1)!$ ways in which they may be arranged. Similarly, the group of values from x_{p+1} to x_{q-1} may be arranged in $(q-p-1)!$ ways and the third group from x_{q+1} to x_n in $(n-q)!$ ways. The distribution of random samples in which the pth and qth values in ascending order are denoted by x_p and x_q, and the remaining values satisfy the conditions in (2), is therefore given by

$$p(x_1, x_2, \ldots, x_n)\,dx_1 dx_2 \ldots dx_n = \left\{\frac{n!}{(p-1)!\,(q-p-1)!\,(n-q)!}\right\}$$
$$\times \frac{1}{(2\pi)^{\frac{1}{2}n}\sigma^n}\exp\left[-\frac{1}{2\sigma^2}\sum_{r=1}^{r=n}(x_r-\xi)^2\right]dx_1 dx_2 \ldots dx_n, \tag{3}$$

where the constant term in brackets makes the total frequency of all such samples equal to unity.

The joint distribution of the sample mean ($\bar{x}$) of the n values and the difference $\Delta = (x_q - x_p)$ is given by

$$p(\bar{x}, \Delta)\,d\bar{x}\,d\Delta = \frac{n!}{(p-1)!\,(q-p-1)!\,(n-q)!}$$
$$\times \frac{1}{(2\pi)^{\frac{1}{2}n}\sigma^n}\int\ldots\int \exp\left[-\frac{1}{2\sigma^2}\sum_{r=1}^{r=n}(x_r-\xi)^2\right]dx_1 dx_2 \ldots dx_n, \tag{4}$$

where the multiple integral is evaluated over the domain of the x's conditioned by the limits indicated in (2) and by $\Delta = (x_q - x_p)$ and $\bar{x} = \frac{1}{n}\sum_{r=1}^{r=n} x_r$.

Make the transformation to variables defined by

$$\left.\begin{aligned}
\bar{x} &= \frac{1}{n}(x_1+x_2+\dots+x_n),\\
y_1 &= -x_p+x_1,\\
y_2 &= -x_p+x_2,\\
&\dots\dots\dots\dots\\
y_{p-1} &= -x_p+x_{p-1},\\
y_{p+1} &= -x_p+x_{p+1},\\
&\dots\dots\dots\dots\\
y_n &= -x_p+x_n.
\end{aligned}\right\} \tag{5}$$

The Jacobian of the transformation is $\dfrac{\partial(x_1\dots x_n)}{\partial(\bar{x}, y_1, \dots, y_{p-1}, y_{p+1}, \dots, y_n)} = 1$, and, from (2) and (5), the transformed limits of integration are

$$\left.\begin{aligned}
-\infty < y_r \leqslant 0, &\quad r = 1, 2, \dots, (p-1),\\
0 \leqslant y_r \leqslant \Delta, &\quad r = (p+1), (p+2), \dots, (q-1),\\
\Delta \leqslant y_r < \infty, &\quad r = (q+1), (q+2), \dots, n.\\
y_q = \Delta. &
\end{aligned}\right\} \tag{6}$$

Using the relations in (5) above it may easily be shown that

$$\sum_{r=1}^{r=n}(x_r-\xi)^2 = n(\bar{x}-\xi)^2+\frac{n-1}{n}\sum_{r=1}^{r=n}y_r^2-\frac{2}{n}\sum_{s=1}^{s=n-1}\sum_{t=1}^{t=n-s}y_s y_{s+t}, \tag{7}$$

where in the summation on the right $r, s, s+t \neq p$. With the new variables we have, from (4), (5), (6) and (7), the joint distribution of $\bar{x}$ and Δ given by

$$p(\bar{x},\Delta)\,d\bar{x}\,d\Delta = \left[\frac{\exp\left[-\dfrac{n(\bar{x}-\xi)^2}{2\sigma^2}\right]}{\sqrt{(2\pi)}\,\sigma/\sqrt{n}}\,d\bar{x}\right]\times\left[\frac{\sqrt{n}\,(n-1)!\,d\Delta}{(p-1)!\,(q-p-1)!\,(n-q)!\,(2\pi)^{\frac{1}{2}(n-1)}\sigma^{n-1}}\right.$$
$$\times\int_{-\infty}^{0}dy_1\dots\int_{-\infty}^{0}dy_{p-1}\int_{0}^{\Delta}dy_{p+1}\dots\int_{0}^{\Delta}dy_{q-1}\int_{\Delta}^{\infty}dy_{q+1}\dots\int_{\Delta}^{\infty}\exp\left[-\frac{1}{2n\sigma^2}\right.$$
$$\left.\left.\times\left\{(n-1)\sum_{r=1}^{r=n}y_r^2-2\sum_{s=1}^{s=n-1}\sum_{t=1}^{t=n-s}y_s y_{s+t}\right\}\right]dy_n\right], \tag{8}$$

with the restriction that $r, s, s+t \neq p$, and Δ is to be substituted for y_q.

The term in the first bracket of (8) is the distribution of the sample mean $\bar{x}$. It follows, therefore, that the term in the second bracket is the distribution of Δ, because this expression does not involve $\bar{x}$ but is a function of Δ alone. This indicates that, in random samples from a normal population, the difference between the pth and qth values in order of magnitude is independent of the sample mean. It follows, therefore, that all estimates of the population standard deviation σ determined from ranked variate differences (e.g. from the semi-interquartile range or other percentile measures of dispersion) are independent of the corresponding sample mean.

As a special case, when $p = 1$ and $q = n$, the difference between the pth and qth values becomes the difference between the lowest and highest, i.e. the range of the sample. Furthermore, if the values in a sample be divided into random subgroups, a simple extension of the argument shows that there is also statistical independence between sample mean and the corresponding mean range of the subgroups.

Table 3. 10 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30	60
2	5·04	2·62	2·20	2·03	1·94	1·89	1·82	1·78	1·73	1·71	1·69	1·67 (1)
3	2·59	2·02	1·88	1·81	1·77	1·75+	1·72	1·71	1·69	1·67	1·66	1·66 (1)
4	2·18	1·88	1·79	1·75+	1·73	1·72	1·70	1·69	1·67	1·67	1·66	1·65+
5	2·02	1·81	1·75+	1·73	1·71	1·70	1·68	1·68	1·67	1·66	1·66	1·65
6	1·94	1·78	1·73	1·71	1·70	1·69	1·68	1·67	1·66	1·66	1·65+	1·65−
7	1·88	1·76	1·72	1·70	1·69	1·68	1·67	1·67	1·66	1·66	1·65+	1·65−
8	1·85	1·74	1·71	1·69	1·68	1·68	1·67	1·66	1·66	1·65+	1·65+	1·65−
9	1·82	1·73	1·70	1·69	1·68	1·67	1·67	1·66	1·66	1·65+	1·65	1·65−
10	1·81	1·72	1·70	1·68	1·68	1·67	1·66	1·66	1·65+	1·65+	1·65	1·65−
12	1·78	1·71	1·69	1·68	1·67	1·67	1·66	1·66	1·65+	1·65+	1·65−	1·65−
14	1·76	1·70	1·68	1·67	1·67	1·66	1·66	1·66	1·65+	1·65	1·65−	1·65−
16	1·75	1·70	1·68	1·67	1·67	1·66	1·66	1·65+	1·65+	1·65	1·65−	1·65−
18	1·74	1·69	1·68	1·67	1·66	1·66	1·66	1·65+	1·65+	1·65−	1·65−	1·65−
20	1·73	1·69	1·67	1·67	1·66	1·66	1·66	1·65+	1·65	1·65−	1·65−	1·65−

Table 4. 5 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30	60
2	10·14	3·87	2·98	2·66	2·49	2·38	2·26	2·20	2·11	2·07	2·03	2·00 (2)
3	3·82	2·64	2·37	2·25	2·19	2·14	2·09	2·07	2·03	2·01	1·99	1·98 (1)
4	2·95+	2·37	2·22	2·15−	2·11	2·08	2·05	2·03	2·01	2·00	1·98	1·97
5	2·63	2·25+	2·15−	2·10	2·07	2·05	2·03	2·01	2·00	1·99	1·98	1·97 (1)
6	2·48	2·19	2·11	2·07	2·05−	2·03	2·01	2·00	1·99	1·98	1·97	1·97 (1)
7	2·38	2·15+	2·09	2·05+	2·03	2·02	2·01	2·00	1·98	1·98	1·97	1·97 (1)
8	2·32	2·13	2·07	2·04	2·02	2·01	2·00	1·99	1·98	1·98	1·97	1·97 (1)
9	2·27	2·11	2·06	2·03	2·02	2·01	2·00	1·99	1·98	1·97	1·97	1·96
10	2·24	2·09	2·05−	2·02	2·01	2·00	1·99	1·98	1·98	1·97	1·97	1·96
12	2·19	2·07	2·03	2·01	2·00	2·00	1·99	1·98	1·97	1·97	1·97	1·96
14	2·16	2·06	2·02	2·01	2·00	1·99	1·98	1·98	1·97	1·97	1·97	1·96
16	2·14	2·05−	2·02	2·00	1·99	1·99	1·98	1·98	1·97	1·97	1·97	1·96
18	2·12	2·04	2·01	2·00	1·99	1·99	1·98	1·98	1·97	1·97	1·97	1·96
20	2·11	2·03	2·01	2·00	1·99	1·98	1·98	1·97	1·97	1·97	1·96	1·96

Note. The numbers in brackets in the column headed $m = 60$ indicate the number of units which must be subtracted in the second decimal place to obtain the level for $m = 120$ and the same value of n. Where no figure is given $u(120, n) = u(60, n)$ to second decimal place accuracy. E.g. for the 5 % level, $u(120, 2) = 1 \cdot 98$.

Table 5. 2 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30	60
2	25·39	6·27	4·27	3·60	3·27	3·08	2·86	2·73	2·59	2·52	2·45+	2·39 (3)
3	6·19	3·56	3·05−	2·84	2·72	2·65−	2·56	2·51	2·45−	2·42	2·39	2·36 (2)
4	4·21	3·05−	2·77	2·65−	2·58	2·53	2·48	2·45−	2·41	2·39	2·37	2·35− (1)
5	3·56	2·84	2·65−	2·56	2·51	2·48	2·44	2·42	2·39	2·37	2·36	2·34 (1)
6	3·25+	2·73	2·58	2·51	2·47	2·45−	2·42	2·40	2·37	2·36	2·35+	2·34 (1)
7	3·07	2·66	2·54	2·48	2·45+	2·43	2·40	2·39	2·37	2·36	2·35−	2·34 (1)
8	2·95+	2·61	2·51	2·46	2·43	2·42	2·39	2·38	2·36	2·35+	2·34	2·34 (1)
9	2·87	2·58	2·49	2·45−	2·42	2·41	2·39	2·37	2·36	2·35+	2·34	2·33
10	2·81	2·55+	2·47	2·44	2·41	2·40	2·38	2·37	2·35+	2·35−	2·34	2·33
12	2·72	2·51	2·45−	2·42	2·40	2·39	2·37	2·36	2·35+	2·34	2·34	2·33
14	2·67	2·49	2·43	2·41	2·39	2·38	2·37	2·36	2·35−	2·34	2·34	2·33
16	2·63	2·47	2·42	2·40	2·38	2·37	2·36	2·35+	2·35−	2·34	2·34	2·33
18	2·60	2·46	2·41	2·39	2·38	2·37	2·36	2·35+	2·34	2·34	2·33	2·33
20	2·58	2·45−	2·41	2·39	2·37	2·37	2·36	2·35−	2·34	2·34	2·33	2·33

Table 6. 1 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30	60
2	50·79	8·93	5·49	4·43	3·93	3·64	3·32	3·14	2·93	2·84	2·75−	2·66 (4)
3	8·82	4·34	3·60	3·30	3·14	3·03	2·91	2·84	2·75−	2·70	2·66	2·62 (2)
4	5·42	3·60	3·20	3·02	2·92	2·86	2·79	2·74	2·68	2·66	2·63	2·60 (1)
5	4·38	3·29	3·02	2·90	2·83	2·79	2·73	2·70	2·66	2·64	2·62	2·60 (1)
6	3·90	3·13	2·93	2·83	2·78	2·74	2·70	2·67	2·64	2·62	2·61	2·59 (1)
7	3·63	3·03	2·87	2·79	2·75−	2·72	2·68	2·66	2·63	2·62	2·60	2·59 (1)
8	3·45+	2·97	2·83	2·76	2·72	2·70	2·67	2·65−	2·62	2·61	2·60	2·59 (1)
9	3·33	2·92	2·80	2·74	2·71	2·68	2·66	2·64	2·62	2·61	2·60	2·59 (1)
10	3·24	2·88	2·78	2·72	2·69	2·67	2·65−	2·63	2·61	2·60	2·59	2·59 (1)
12	3·12	2·83	2·74	2·70	2·68	2·66	2·64	2·62	2·61	2·60	2·59	2·58
14	3·05−	2·80	2·72	2·69	2·66	2·65−	2·63	2·62	2·60	2·60	2·59	2·58
16	2·99	2·78	2·71	2·67	2·65+	2·64	2·62	2·61	2·60	2·60	2·59	2·58
18	2·95−	2·76	2·70	2·66	2·65−	2·63	2·62	2·61	2·60	2·59	2·59	2·58
20	2·92	2·74	2·69	2·66	2·64	2·63	2·62	2·61	2·60	2·59	2·59	2·58

Note. The numbers in brackets in the column headed $m = 60$ indicate the number of units which must be subtracted in the second decimal place to obtain the level for $m = 120$ and the same value of n. Where no figure is given $u(120, n) = u(60, n)$ to second decimal place accuracy. E.g. for the 2 % level $u(120, 5) = 2·33$.

Table 7. 0·2 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30
2	254·0	20·1	9·4	6·8	5·7	5·1	4·5−	4·1	3·7	3·6	3·4 (2)
3	19·8	6·8	5·1	4·5	4·2	4·0	3·7	3·6	3·4	3·3	3·2
4	9·4	5·1	4·2	3·9	3·7	3·6	3·5−	3·4	3·3	3·2	3·2 (1)
5	6·8	4·4	3·9	3·7	3·5+	3·5−	3·4	3·3	3·2	3·2	3·2 (1)
6	5·7	4·1	3·7	3·5+	3·4	3·4	3·3	3·3	3·2	3·2	3·1
7	5·1	3·9	3·6	3·5−	3·4	3·3	3·3	3·2	3·2	3·2	3·1
8	4·7	3·8	3·5+	3·4	3·3	3·3	3·2	3·2	3·2	3·2	3·1
9	4·5	3·7	3·5−	3·4	3·3	3·3	3·2	3·2	3·2	3·1	3·1
10	4·3	3·6	3·4	3·3	3·3	3·3	3·2	3·2	3·2	3·1	3·1
12	4·1	3·5+	3·4	3·3	3·3	3·2	3·2	3·2	3·1	3·1	3·1
15	3·9	3·5−	3·3	3·3	3·2	3·2	3·2	3·2	3·1	3·1	3·1
20	3·7	3·4	3·3	3·2	3·2	3·2	3·2	3·1	3·1	3·1	3·1

Table 8. 0·1 % *points of* $u = u(m, n)$

n \ m	1	2	3	4	5	6	8	10	15	20	30
2	507·9	28·5−	11·7	8·1	6·7	5·9	5·0	4·6	4·1	3·9	3·7 (2)
3	28·1	8·7	6·0	5·1	4·6	4·3	4·0	3·9	3·7	3·6	3·5 (1)
4	11·8	5·8	4·7	4·3	4·1	3·9	3·7	3·6	3·5+	3·5−	3·4 (1)
5	8·2	5·0	4·3	4·0	3·8	3·7	3·6	3·6	3·5−	3·4	3·4 (1)
6	6·7	4·5+	4·0	3·8	3·7	3·6	3·5+	3·5−	3·4	3·4	3·4 (1)
7	5·9	4·3	3·9	3·7	3·6	3·6	3·5+	3·5−	3·4	3·4	3·3
8	5·4	4·1	3·8	3·7	3·6	3·5+	3·5−	3·4	3·4	3·4	3·3
9	5·0	4·0	3·8	3·6	3·6	3·5+	3·5−	3·4	3·4	3·4	3·3
10	4·8	4·0	3·7	3·6	3·5+	3·5	3·4	3·4	3·4	3·4	3·3
12	4·5+	3·8	3·6	3·6	3·5	3·5−	3·4	3·4	3·4	3·3	3·3
15	4·2	3·7	3·6	3·5+	3·5−	3·4	3·4	3·4	3·3	3·3	3·3
20	4·0	3·6	3·5+	3·5−	3·4	3·4	3·4	3·4	3·3	3·3	3·3

Note. The numbers in brackets in the column headed $m = 30$ indicate the number of units which must be subtracted in the first decimal place to obtain the level for $m = 60$ and the same value of n. Where no figure is given $u(60, n) = u(30, n)$ to first decimal place accuracy; $u(120, n) = u(60, n)$ for (i) all 0·2 % points except that $u(120, 3) = 3{\cdot}1$ and (ii) all 0·1 % points except that $u(120, 2) = 3{\cdot}4$ and $u(120, 3) = 3{\cdot}3$.

Table 9. *Table for testing the significance of the deviation of the mean of a small sample (of size n) from some pre-assigned value*

n \ α	0·10	0·05	0·02	0·01	0·002	0·001
2	3·157	6·353	15·910	31·828	159·16	318·31
3	0·885−	1·304	2·111	3·008	6·77	9·58
4	·529	0·717	1·023	1·316	2·29	2·85+
5	·388	·507	0·685+	0·843	1·32	1·58
6	0·312	0·399	0·523	0·628	0·92	1·07
7	·263	·333	·429	·507	·71	0·82
8	·230	·288	·366	·429	·59	·67
9	·205−	·255+	·322	·374	·50	·57
10	·186	·230	·288	·333	·44	·50
11	0·170	0·210	0·262	0·302	0·40	0·44
12	·158	·194	·241	·277	·36	·40
13	·147	·181	·224	·256	·33	·37
14	·138	·170	·209	·239	·31	·34
15	·131	·160	·197	·224	·29	·32
16	0·124	0·151	0·186	0·212	0·27	0·30
17	·118	·144	·177	·201	·26	·28
18	·113	·137	·168	·191	·24	·26
19	·108	·131	·161	·182	·23	·25+
20	·104	·126	·154	·175−	·22	·24

The table gives values of the ratio $\frac{\delta}{w} = \frac{\text{deviation of sample mean}}{\text{range in sample}}$ lying on different levels of significance, the levels being the sum, α, of the two tails of the probability distribution.

Table 10. *Table for testing the significance of the difference between the means of two small samples of equal size n*

n \ α	0·10	0·05	0·02	0·01	0·002	0·001
2	2·322	3·427	5·553	7·916	17·81	25·23
3	0·974	1·272	1·715−	2·093	3·27	4·18
4	·644	0·813	1·047	1·237	1·74	1·99
5	·493	·613	0·772	0·896	1·21	1·35+
6	0·405+	0·499	0·621	0·714	0·94	1·03
7	·347	·426	·525+	·600	·77	0·85−
8	·306	·373	·459	·521	·67	·73
9	·275−	·334	·409	·464	·59	·64
10	·250	·304	·371	·419	·53	·58
11	0·233	0·280	0·340	0·384	0·48	0·52
12	·214	·260	·315+	·355+	·44	·48
13	·201	·243	·294	·331	·41	·45−
14	·189	·228	·276	·311	·39	·42
15	·179	·216	·261	·293	·36	·39
16	0·170	0·205−	0·247	0·278	0·34	0·37
17	·162	·195+	·236	·264	·33	·35+
18	·155+	·187	·225+	·252	·31	·34
19	·149	·179	·216	·242	·30	·32
20	·143	·172	·207	·232	·29	·31

The table gives values of the ratio $\frac{|\bar{x}_1 - \bar{x}_2|}{\frac{1}{2}(w' + w'')} = \frac{\text{difference between means}}{\text{mean of sample ranges}}$ lying on different levels of significance. The levels are the sum, α, of the two tails of the probability distribution.

N.B. When considering deviations in the positive (or negative) direction only, the values of α at the headings of the columns should be halved.

Table 11

n	d_n	$1/d_n$	$\sqrt{n}$	$d_n\sqrt{n}$
2	1·1284	0·8862	1·4142	1·5958
3	1·6926	·5908	1·7321	2·9316
4	2·0588	·4857	2·0000	4·1175
5	2·3259	·4299	2·2361	5·2009
6	2·5344	0·3946	2·4495	6·2080
7	2·7044	·3698	2·6458	7·1551
8	2·8472	·3512	2·8284	8·0531
9	2·9700	·3367	3·0000	8·9101
10	3·0775	·3249	3·1623	9·7319
11	3·1729	0·3152	3·3166	10·5232
12	3·2585	·3069	3·4641	11·2876
13	3·3360	·2998	3·6056	12·0281
14	3·4068	·2935	3·7417	12·7469
15	3·4718	·2880	3·8730	13·4463
16	3·5320	0·2831	4·0000	14·1279
17	3·5879	·2787	4·1231	14·7932
18	3·6401	·2747	4·2426	15·4435
19	3·6890	·2711	4·3589	16·0798
20	3·7350	·2677	4·4721	16·7032

REFERENCES

Daly, J. F. (1946). *Ann. Math. Statist.* **17**, 71.

Davies, O. L. & Pearson, E. S. (1934). *J. Roy. Statist. Soc. Suppl.* **1**, 76.

Fisher, R. A. (1925). *Metron*, **5**, 90.

Hartley, H. O. (1942). *Biometrika*, **32**, 334.

McKay, A. T. & Pearson, E. S. (1933). *Biometrika*, **25**, 415.

Pearson, E. S. (1926). *Biometrika*, **18**, 173.

Pearson, E. S. (1932). *Biometrika*, **24**, 404.

Pearson, E. S. (1935). *The Application of Statistical Methods to Industrial Standardization and Quality Control.* B.S. No. 600. London: British Standards Institution.

Pearson, E. S. & Haines, Joan (1935). *J. Roy. Statist. Soc. Suppl.* **2**, 83.

Pearson, E. S. & Hartley, H. O. (1942). *Biometrika*, **32**, 301.

'Student' (1908). *Biometrika*, **6**, 1.

Tippett, L. H. C. (1925). *Biometrika*, **17**, 364.

THE FREQUENCY DISTRIBUTION OF $\sqrt{b_1}$ FOR SAMPLES OF ALL SIZES DRAWN AT RANDOM FROM A NORMAL POPULATION

By R. C. GEARY

1. Introductory

A research on which the writer has been engaged for some years has so far yielded the following results:

(1) Testing for normality has a greater practical importance than statisticians (including the writer) have been disposed to accord to it; actual probabilities may be seriously at variance with probabilities derived from the well-known tables computed on the hypothesis of universal normality; in consequence, testing for normality and, where necessary, correction (even if rough and tentative) for suspected universal non-normality, should become a part of statistical routine.

(2) For large samples, $\sqrt{b_1}$ and b_2 are the most efficient of large fields of tests of skewness and kurtosis, respectively, amongst large fields of alternative universes.

These matters will be dealt with in detail in subsequent papers. It seems, in the first instance, desirable to derive the frequency distribution of $\sqrt{b_1}$ for normal random samples of *all* sizes, partly on account of the inherent importance of the problem, partly in order to explore a computational technique which might be found effective in solving the analogous but probably more difficult b_2 problem.

Towards the solution of the problem there are available the exact values of first four even moments—the odd moments are, of course, zero—of normal $\sqrt{b_1}$, the second, *fourth* and sixth having been determined by R. A. Fisher (1930) and the eighth by Joseph Pepper (1932). It may be useful here to set out the four moments. Taking

$$\sqrt{b_1} = m_3/m_2^{\frac{3}{2}} = n^{\frac{1}{2}}\left\{\sum_{i=1}^{n}(x_i-\bar{x})^3\right\}\Big/\{\Sigma(x_i-\bar{x})^2\}^{\frac{3}{2}}, \tag{1·1}$$

where n is the sample number, we have

$$\left.\begin{aligned}
\mu_2 &= \frac{6(n-2)}{(n+1)(n+3)},\\
\mu_4 &= \frac{108(n-2)(n^2+27n-70)}{(n+1)(n+3)(n+5)(n+7)(n+9)},\\
\mu_6 &= \frac{3240(n-2)(n^4+84n^3+2695n^2-15168n+20020)}{(n+1)(n+3)(n+5)(n+7)(n+9)(n+11)(n+13)(n+15)},\\
\mu_8 &= \frac{7.5.3^5.2^4(n-2)(n^6+171n^5+13893n^4+580401n^3-5131014n^2 + 14132268n-12932920)}{(n+1)(n+3)(n+5)\dots(n+17)(n+19)(n+21)}.
\end{aligned}\right\} \tag{1·2}$$

E. S. Pearson (1931, 1936) derived empirically 0·05 and 0·01 probability points for certain values of $n \geqslant 25$ using a Pearson Type VII curve and earlier approximations by R. A. Fisher (1929) of the second and fourth moments.

The method here used for the derivation of the frequency distribution of $\sqrt{b_1}$ is essentially an elaboration of that which the author used (1935, 1936) for finding the frequency distribution of the test of kurtosis a (the ratio of the mean deviation to the standard deviation of the numbers sampled), which consisted in establishing a relation in integral form between the frequency ordinate for n with the value for $(n-1)$ and thereby determining the ordinates to any required degree of accuracy for the lower n's. At a certain stage the actual frequency is shown to be very close to the value based on the Gram-Charlier curve for the same value of n; and the assumption is made that the Gram-Charlier may be relied on for values of n greater than the 'transition value'. In the present problem the known normal moments are utilized as well at every stage. In the concluding section the status of the solution in the hierarchy of 'precision' is discussed.

Since the frequency is symmetrical, attention is confined practically exclusively to the positive sector.

2. The general integral iteration

To distinguish the sample size by the notation let the value of $\sqrt{b_1}$ be indicated by t_n. Apply a Helmert orthogonal transformation to the original observations $x_1, x_2, \ldots, x_n$ so that

$$\left.\begin{aligned} x_1' &= (x_1-x_2)/\sqrt{2}, \\ x_2' &= (x_1+x_2-2x_3)/\sqrt{6}, \\ &\vdots \\ x_{n-1}' &= (x_1+x_2+\ldots+x_{n-1}-\overline{n-1}\,x_n)/\sqrt{[n(n-1)]}, \\ x_n' &= (x_1+x_2+\ldots+x_n)/\sqrt{n} = \bar{x}\sqrt{n}, \end{aligned}\right\} \tag{2·1}$$

which, on inversion, gives

$$\left.\begin{aligned} x_1-\bar{x} &= \frac{x_1'}{\sqrt{2}}+\frac{x_2'}{\sqrt{6}}+\ldots+\frac{x_{n-1}'}{\sqrt{[n(n-1)]}}, \\ x_2-\bar{x} &= -\frac{x_1'}{\sqrt{2}}+\frac{x_2'}{\sqrt{6}}+\ldots+\frac{x_{n-1}'}{\sqrt{[n(n-1)]}}, \\ x_3-\bar{x} &= -\frac{2x_2'}{\sqrt{6}}+\ldots+\frac{x_{n-1}'}{\sqrt{[n(n-1)]}}, \\ &\vdots \\ x_{n-1}-\bar{x} &= -\frac{(n-2)\,x_{n-2}'}{\sqrt{[(n-1)(n-2)]}}+\frac{x_{n-1}'}{\sqrt{[n(n-1)]}}, \\ x_n-\bar{x} &= -\frac{(n-1)\,x_{n-1}'}{\sqrt{[n(n-1)]}}. \end{aligned}\right\} \tag{2·2}$$

Then

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n-1} x_i'^2, \tag{2·3}$$

and

$$\left.\begin{aligned}\Sigma(x_i-\bar{x})^3 = {} & 3x_1'^2\left(\frac{x_2'}{\sqrt{6}}+\frac{x_3'}{\sqrt{12}}+\ldots+\frac{x_{n-1}'}{\sqrt{[(n-1)n]}}\right)\\ & +3x_2'^2\left(\frac{x_3'}{\sqrt{12}}+\frac{x_4'}{\sqrt{20}}+\ldots+\frac{x_{n-1}'}{\sqrt{[(n-1)n]}}\right)\\ & +3x_3'^2\left(\frac{x_4'}{\sqrt{20}}+\frac{x_5'}{\sqrt{30}}+\ldots+\frac{x_{n-1}'}{\sqrt{[(n-1)n]}}\right)\\ & \qquad\vdots\\ & \qquad\qquad +\frac{3x_{n-2}'^2x_{n-1}'}{\sqrt{[(n-1)n]}}\\ & -\frac{x_2'^3}{\sqrt{6}}-\frac{2x_3'^3}{\sqrt{12}}-\frac{3x_4'^3}{\sqrt{20}}-\ldots-\frac{(n-2)x_{n-1}'^3}{\sqrt{[(n-1)n]}}.\end{aligned}\right\} \tag{2·4}$$

Apply a polar transformation to the x_i', that is,

$$\left.\begin{aligned} x_1' &= r\sin\phi_{n-3}\sin\phi_{n-4}\ldots\sin\phi_1\sin\phi_0,\\ x_2' &= r\sin\phi_{n-3}\sin\phi_{n-4}\ldots\sin\phi_1\cos\phi_0,\\ x_3' &= r\sin\phi_{n-3}\sin\phi_{n-4}\ldots\sin\phi_2\cos\phi_1,\\ x_4' &= r\sin\phi_{n-3}\sin\phi_{n-4}\ldots\sin\phi_3\cos\phi_2,\\ &\vdots\\ x_{n-3}' &= r\sin\phi_{n-3}\sin\phi_{n-4}\cos\phi_{n-5},\\ x_{n-2}' &= r\sin\phi_{n-3}\cos\phi_{n-4},\\ x_{n-1}' &= r\cos\phi_{n-3},\end{aligned}\right\} \tag{2·5}$$

and

$$\Sigma x_i'^2 = \Sigma(x_i-\bar{x})^2 = r^2, \tag{2·6}$$

$$\begin{aligned} t_n = n^{\frac{1}{2}}\Big\{ & \frac{3}{\sqrt{6}}\sin^3\phi_{n-3}\sin^3\phi_{n-4}\ldots\sin^3\phi_1\sin^2\phi_0\cos\phi_0\\ & +\frac{3}{\sqrt{12}}\sin^3\phi_{n-3}\sin^3\phi_{n-4}\ldots\sin^3\phi_2\sin^2\phi_1\cos\phi_1\\ & +\ldots+\frac{3}{\sqrt{[(n-2)(n-1)]}}\sin^3\phi_{n-3}\sin^2\phi_{n-4}\cos\phi_{n-4}+\frac{3}{\sqrt{[(n-1)n]}}\sin^2\phi_{n-3}\cos\phi_{n-3}\\ & -\frac{1}{\sqrt{6}}\sin^3\phi_{n-3}\sin^3\phi_{n-4}\ldots\cos^3\phi_0-\frac{2}{\sqrt{12}}\sin^3\phi_{n-3}\ldots\sin^3\phi_2\cos^3\phi_1\\ & -\ldots-\frac{(n-3)}{\sqrt{[(n-2)(n-1)]}}\sin^3\phi_{n-3}\cos^3\phi_{n-4}-\frac{(n-2)}{\sqrt{[(n-1)n]}}\cos^3\phi_{n-3}\Big\},\end{aligned} \tag{2·7}$$

whence the fundamental iteration

$$\frac{t_n}{n^{\frac{1}{2}}} = \frac{t_{n-1}}{(n-1)^{\frac{1}{2}}}\sin^3\phi_{n-3}+\frac{3}{[(n-1)n]^{\frac{1}{2}}}\sin^2\phi_{n-3}\cos\phi_{n-3}-\frac{(n-2)}{[(n-1)n]^{\frac{1}{2}}}\cos^3\phi_{n-3}, \tag{2·8}$$

in which there intervenes only the angle ϕ_{n-3}; and for normal random samples it is a well-known fact that the ϕ_i are distributed independently of one another, the distribution of ϕ_{n-3} being of the form

$$C\sin^{n-3}\phi_{n-3}d\phi_{n-3}. \tag{2·9}$$

Now t_{n-1} involves only $\phi_0, \ldots, \phi_{n-4}$; hence it is independent of ϕ_{n-3}. Accordingly, if the frequency distribution of t_{n-1} is of the form

$$f_{n-1}(t_{n-1})\,dt_{n-1}, \tag{2·10}$$

the joint distribution of ϕ_{n-3} and t_{n-1} is given by

$$C \sin^{n-3}\phi_{n-3}\, d\phi_{n-3} \times f_{n-1}(t_{n-1})\, dt_{n-1}. \tag{2·11}$$

Now, from (2·8),

$$dt_{n-1} = dt_n \left(\frac{n-1}{n}\right)^{\frac{1}{2}} \sin^{-3}\phi_{n-3}. \tag{2·12}$$

On substituting in (2·11) and integrating we find for frequency of t_n the expression

$$f_n(t_n) = \left(\frac{n-1}{\pi n}\right)^{\frac{1}{2}} \frac{\frac{1}{2}(n-3)\,!}{\frac{1}{2}(n-4)\,!} \int d\phi_{n-3} \sin^{n-6}\phi_{n-3} f_{n-1}(t_{n-1}), \tag{2·13}$$

where the relation (2·8) obtains. Integration extends to values of ϕ_{n-3} (so that $0 \leqslant \phi_{n-3} \leqslant \pi$ for $n > 3$) which yield non-zero values of f_{n-1}. Setting $\cos\phi_{n-3} = x$ the integral at (2·13) assumes the form

$$f_n(t_n) = \left(\frac{n-1}{\pi n}\right)^{\frac{1}{2}} \frac{\frac{1}{2}(n-3)\,!}{\frac{1}{2}(n-4)\,!} \int dx (1-x^2)^{\frac{1}{2}(n-7)} f_{n-1}(t_{n-1}), \tag{2·14}$$

with, from (2·8),

$$n^{\frac{1}{2}} t_{n-1} = [(n-1)^{\frac{1}{2}} t_n - 3x + (n+1)\,x^3]/(1-x^2)^{\frac{3}{2}}. \tag{2·15}$$

In the derivation of the frequencies for $n = 4$ to 8 inclusive, dealt with in later sections, both the forms (2·13) and (2·14) are used.

3. FUNCTIONAL DISCONTINUITIES OF THE FREQUENCY

In the integral at (2·14) t_n appears merely as a parameter. Consequently the nature of the frequency $f_n(t_n)$ depends to a considerable extent on the simple algebraic properties of $t_{n-1}(x)$ given by (2·15). The following property (easily demonstrated) is fundamental:

For

$$t_n = (n-2k)/[k(n-k)]^{\frac{1}{2}} = {}_k\tau_n \quad (k = 1, 2, \ldots), \tag{3·1}$$

$t_{n-1}(x)$ has a maximum value of

$$(n-2k+1)/[(k-1)(n-k)]^{\frac{1}{2}} = {}_{k-1}\tau_{n-1} \tag{3·2}$$

for

$$x = -[(n-k)/k(n-1)]^{\frac{1}{2}} = {}_k\xi'_n, \tag{3·3}$$

and a minimum value of

$$(n-2k-1)/[k(n-k-1)]^{\frac{1}{2}} = {}_k\tau_{n-1} \tag{3·4}$$

for

$$x = [k/(n-1)(n-k)]^{\frac{1}{2}} = {}_k\xi''_n. \tag{3·5}$$

DEFINITION. ${}_k\tau_n$ are termed the *link values* or *links* of t_n. The regions between consecutive links are termed *zones*. The graph of $t_{n-1}(x)$ for $-1 \leqslant x \leqslant +1$ and $t_n = {}_k\tau_n$ (given at (3·1)) is illustrated in Fig. 1. The limits of integration for integral (2·14) are now seen to be ${}_k\lambda'_n$ and ${}_k\lambda''_n$ which are the values of x at which the ordinates of the curve (2·15) in (x, t_{n-1}), with parameter $t_n = {}_k\tau_n$, assume the limiting values $+{}_1\tau_{n-1}$ and $-{}_1\tau_{n-1}$. The scale on the right shows the links of t_{n-1}. The curve $t_n = {}_k\tau_n$ traverses all the zones but has a 'turn' in the $(k-1)$th zone, remaining entirely in the zone the while. It is due to this turn that the phenomenon of functional discontinuity manifests itself in the frequency $f_n(t_n)$.

Assume that within the kth zone the frequency $f_{n-1}(t_{n-1})$ is represented by ${}_kf_{n-1}(t_{n-1})$, different in functional form for different values of k but the same (for example, having the same coefficients in a power series) within each zone. It will at once be evident, from (2·14), that the frequency of t_n will have a like property. Now, from (2·4) and (2·5) it will be seen that

$$t_3 = -\frac{\cos 3\phi_0}{\sqrt{2}}; \tag{3·6}$$

the distribution of ϕ_0 is rectangular, so that the distribution* of t_3 is given by

$$f_3(t_3) = \frac{\sqrt{2}}{\pi\sqrt{(1-2t_3^2)}}, \quad |t_3| \leqslant 1/\sqrt{2} \tag{3·7}$$

and zero for $|t_3| > 1/\sqrt{2}$. It follows that t_3 has a functional discontinuity at its links $\pm 1/\sqrt{2}$. *Hence, by iteration, the frequency of t_n is represented by different functional expressions in its different interlink zones.*

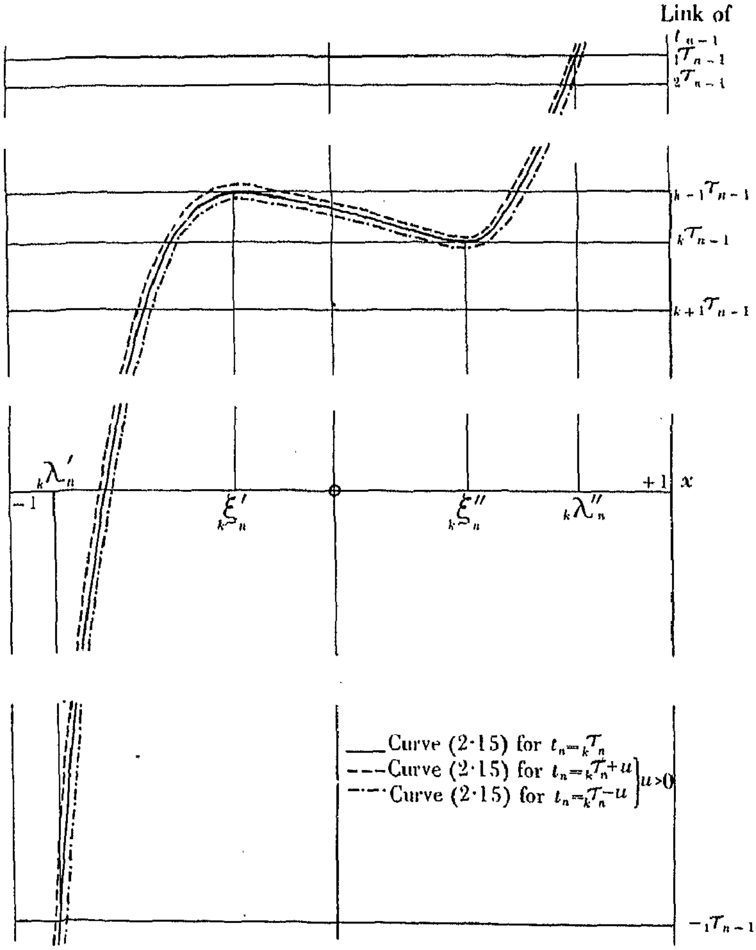


Fig. 1. Graph of $t_{n-1}(x)$.

That the frequency has a finite limit ${}_1\tau_n$ (when n is finite) is established as follows. It can easily be seen from (2·15) that when $t_n = {}_1\tau_n$ the curve $t_{n-1} = t_{n-1}(x)$ degenerates into (i) the straight line $x = -1$ and (ii) a section above the straight line $t_{n-1} = {}_1\tau_{n-1}$ but touching it. For $t_n > {}_1\tau_n$ no part of the curve $t_{n-1} = t_{n-1}(x)$ falls within the rectangle $x = \pm 1$, $t_{n-1} = \pm {}_1\tau_{n-1}$. Reference to (2·14) shows at once that, if $f_{n-1}(t_{n-1}) = 0$ for $|t_{n-1}| > {}_1\tau_{n-1}$, then $f_n(t_n) = 0$ for $|t_n| > {}_1\tau_n$. But (3·7) shows that t_3 has as limiting values $\pm 1/\sqrt{2}$. Hence, by iteration, it follows that the limiting values of the frequency of t_n (or *simpliciter* of t_n) are

$$\pm {}_1\tau_{n-1} = \pm (n-2)/(n-1)^{\frac{1}{2}}. \tag{3·8}$$

* R. A. Fisher (1930).

As will presently appear, the frequencies for $n = 4$ and 5 have marked irregularities: successive integration in accordance with (2·14) imparts, of course, a progressively increasing degree of smoothness to the frequency. To give mathematical expression to this feature, recourse is had to the idea of *order of contact*.

DEFINITION. *Two functions are said to have contact of order* ${}_k\gamma_n$ *at link* ${}_k\tau_n$ *if the functions and their first* $({}_k\gamma_n - 1)$ *derivatives are finite and equal at the link.* It can be shown without difficulty that

$$ {}_k\gamma_n = {}_{k-1}\gamma_{n-1} + 1, \tag{3·9} $$

when $k > 1$, $n > 4$. For what follows it will be convenient to set out for the smaller sample numbers the values of the links and their orders of contact. The links for positive values only of the variables are shown. The orders of contact ${}_1\gamma_n$ will appear from a proposition proved in § 5, giving the actual values of the frequencies near the limit of range. The non-diminishing smoothness in the direction of the centre of the range will be noted.

Values of ${}_k\tau_n$ *and* ${}_k\gamma_n$ *for* $n = 3$ *to* 8 *inclusive*

n	1st link		2nd link		3rd link		4th link	
	${}_1\tau_n$	${}_1\gamma_n$	${}_2\tau_n$	${}_2\gamma_n$	${}_3\tau_n$	${}_3\gamma_n$	${}_4\tau_n$	${}_4\gamma_n$
3	$1/\sqrt{2}$	0	—	—	—	—	—	—
4	$2/\sqrt{3}$	0	0	0	—	—	—	—
5	$3/2$	1	$1/\sqrt{6}$	1	—	—	—	—
6	$4/\sqrt{5}$	1	$1/\sqrt{2}$	2	0	2	—	—
7	$5/\sqrt{6}$	2	$3/\sqrt{10}$	2	$1/2\sqrt{3}$	3	—	—
8	$6/\sqrt{7}$	2	$2/\sqrt{3}$	3	$2/\sqrt{15}$	3	0	4

For even values of n the origin is always a link. In the determination of the frequencies for $n = 5$ to 8, by the methods described in subsequent sections, the link ordinates and the central ordinate play a cardinal role. In fact, the method will be seen to consist essentially in finding curves which pass through the central and link ordinal points, have the required orders of contact and the required form at the limit of range and have the exact earlier momental values (see first section).

4. THE FREQUENCY NEAR THE CENTRE OF RANGE

It will first be shown that

$$ f_n'(+0) = 0 \quad \text{for} \quad n > 4. \tag{4·1} $$

In fact, from (2·14) and (2·15) if $t_n = u$, a small positive quantity,

$$ f_n(u) = C\int_{-\lambda-\kappa u=\lambda'}^{+\lambda-\kappa u=\lambda''} dx(1-x^2)^{\frac{1}{2}(n-7)} f_{n-1}[t_{n-1}(x)], $$

λ, κ being positive constants. Hence

$$ f_n'(u) = -C\kappa\{(1-\lambda''^2)^{\frac{1}{2}(n-7)} f_{n-1}[t_{n-1}(\lambda'')] - (1-\lambda'^2)^{\frac{1}{2}(n-7)} f_{n-1}[t_{n-1}(\lambda')]\} $$

$$ + C'\int_{\lambda'}^{\lambda''} dx(1-x^2)^{\frac{1}{2}(n-10)} f_{n-1}'[t_{n-1}(x)]. $$

Letting $u \to 0$ the integral-free expression obviously vanishes provided that $f_{n-1}[t_{n-1}(\lambda)]$ is finite, which it is when $n > 4$; and the integral becomes

$$\int_{-\lambda}^{+\lambda} dx(1-x^2)^{\frac{1}{2}(n-10)} f'_{n-1}\left(\frac{\overline{n+1}\,x^3 - 3x}{\overline{1-x^2}^{\frac{3}{2}}}\right).$$

Since $f_{n-1}(y)$ is an even function of y, its derivative is odd which remains an odd function when y is replaced by an odd function of x. Hence the integral vanishes.

5. The form of the frequency at the limit of range

In this section it will be shown that *near* $t_n = \pm(n-2)/(n-1)^{\frac{1}{2}}$ *the frequency is given by*

$$f_n(t_n) = \frac{1}{3\sqrt{\pi}} \frac{\frac{1}{2}(n-3)!}{\frac{1}{2}(n-4)!} \frac{(n-1)^{\frac{1}{2}(n-3)}}{(3n.\overline{n-2})^{\frac{1}{2}(n-4)}} \left(\frac{\overline{n-2}^2}{n-1} - t_n^2\right)^{\frac{1}{2}(n-4)}. \tag{5·1}$$

It may be seen at once that for $n = 3$ the frequency by (5·1) would be

$$f_3(t_3) = \frac{1}{\pi}(\tfrac{1}{2} - t_3^2)^{-\frac{1}{2}}, \tag{5·2}$$

as at (3·7). For $n = 4$, (5·1) gives $\frac{1}{2}\sqrt{3}$, which is the value found by A. T. McKay (1933).

The general theorem will be proved by iteration. We assume a general form

$$f_{n-1}(t_{n-1}) = C_{n-1}\left(\frac{\overline{n-3}^2}{n-2} - t_{n-1}^2\right)^{\frac{1}{2}(n-5)}, \tag{5·3}$$

and show that a similar form emerges for $f_n(t_n)$, finding incidentally an iteration relation for the constant C_n. First set

$$v = \overline{n-2} - \overline{n-1}^{\frac{1}{2}}\, t_n,$$

and assume that v is a positive quantity. It will readily appear, from (2·15), that, for $v = 0$, $t_{n-1}(x)$ has a double root at $x = 1/(n-1)$. Accordingly we set

$$x = x' + 1/(n-1) \tag{5·4}$$

and

$$X = \frac{(n-3)^2}{(n-2)} - t_{n-1}^2. \tag{5·5}$$

Having regard only to principal terms we find

$$1 - x^2 \simeq^* \frac{n(n-2)}{(n-1)^2}, \tag{5·6}$$

$$X \simeq 2n^{-2}(n-1)^3(n-2)^{-2}(n-3)(1-x''^2)\,v^{\frac{1}{2}}, \tag{5·7}$$

with

$$x' = \left(\frac{2}{3}\frac{\overline{n-2}}{\overline{n-1}^2}\right)^{\frac{1}{2}} v^{\frac{1}{2}} x''. \tag{5·8}$$

Now, from (2·14),

$$f_n(t_n) = \left(\frac{n-1}{\pi n}\right)^{\frac{1}{2}} \frac{\frac{1}{2}(n-3)!}{\frac{1}{2}(n-4)!} \int_D dx(1-x^2)^{\frac{1}{2}(n-7)} f_{n-1}(t_{n-1}),$$

and, from the analysis in § 3, it will be clear that there are two separate parts of the domain D:

(I) a part near $x = 1/(n-1)$ for which t_{n-1} is entirely in the first zone and by hypothesis has the form (5·3);

(II) a part near $x = -1$ in which t_{n-1} assumes all values.

* The symbol $\simeq$ signifies 'equals, to required approximation'.

Let
$$f_n(t_n) = f_n^{\mathrm{I}}(t_n) + f_n^{\mathrm{II}}(t_n), \tag{5·9}$$
where the functions on the right represent the contributions accruing from the respective parts of the domain of integration. Then

$$f_n^{\mathrm{I}}(t_n) \simeq \frac{\frac{1}{2}(n-3)!}{\frac{1}{2}(n-4)!\sqrt{\pi}}\left(\frac{n-1}{n}\right)^{\frac{1}{2}} \int_{x''=-1}^{x''=+1} dx(1-x^2)^{\frac{1}{2}(n-7)}\, C_{n-1}\{2n^{-2}(n-1)^3 \times (n-2)^{-2}\,(n-3)\,(1-x''^2)\, v^{\frac{1}{2}}\}^{\frac{1}{2}(n-5)},$$

which, on a change of variable from x to x'' by (5·4) and (5·8) and integrating in x'', becomes

$$\simeq C_{n-1}\, 3^{-\frac{1}{2}} \frac{\frac{1}{2}(n-3)!\,\frac{1}{2}(n-5)!}{\frac{1}{2}(n-4)!^2} n^{-\frac{1}{2}(n-2)}\,(n-1)^{n-3}\,(n-2)^{-(n-4)}\,(n-3)^{\frac{1}{2}(n-5)} \left(\frac{\overline{n-2}^2}{n-1} - t_n^2\right)^{\frac{1}{2}(n-4)}, \tag{5·10}$$

which is of the required form. As regards the contribution of II, in (2·15) set

$$x+1 = \beta v + y v^{\frac{3}{2}}. \tag{5·11}$$

It will be found that when t_{n-1} has its limiting value $+(n-3)/(n-2)^{\frac{1}{2}}$ the vanishing of the terms in v^2 and v^3 gives

$$\beta = \tfrac{1}{3}n \quad \text{and} \quad y = \frac{2}{9}\sqrt{\frac{2}{3}\frac{(n-3)}{n^2(n-2)^{\frac{1}{2}}}},$$

whereas if t_{n-1} has its limiting value $-(n-3)/(n-2)^{\frac{1}{2}}$ the values are

$$\beta = \tfrac{1}{3}n \quad \text{and} \quad y = -\frac{2}{9}\sqrt{\frac{2}{3}\frac{(n-3)}{n^2(n-2)^{\frac{1}{2}}}}.$$

Between the limits of x,

$$-1 + \frac{v}{3n} \pm \frac{2}{9}\sqrt{\frac{2}{3}\frac{n-3}{n^2(n-2)^{\frac{1}{2}}}}\, v^{\frac{3}{2}}, \tag{5·12}$$

$t_{n-1}(x)$ given by (2·15), assumes once all values between its limits of range and, in fact,

$$y \simeq \frac{2}{9}\sqrt{\frac{2}{3}\frac{t_{n-1}}{n^2}}. \tag{5·13}$$

Now

$$f_n^{\mathrm{II}}(t_n) = \frac{\left(\frac{n-1}{n}\right)^{\frac{1}{2}} \frac{1}{2}(n-3)!}{\sqrt{\pi}\,\frac{1}{2}(n-4)!} \int dx(1-x^2)^{\frac{1}{2}(n-7)} f_{n-1}(t_{n-1}), \tag{5·14}$$

the limits of integration in x being given by (5·12). By (5·11) and (5·13) change the variable x into t_{n-1} (via y) when (5·14) becomes

$$f_n^{\mathrm{II}}(t_n) \simeq C(n)\, v^{\frac{1}{2}(n-4)} \int f_{n-1}(t_{n-1})\, dt_{n-1}, \tag{5·15}$$

and the integral on the right is unity. Written in full (5·15) then becomes

$$f_n^{\mathrm{II}}(t_n) \simeq \frac{1}{3\sqrt{\pi}\, n} \frac{\frac{1}{2}(n-3)!}{\frac{1}{2}(n-4)!} \frac{(n-1)^{\frac{1}{2}(n-3)}}{(3n\,\overline{n-2})^{\frac{1}{2}(n-4)}} \left(\frac{\overline{n-2}^2}{n-1} - t_n^2\right)^{\frac{1}{2}(n-4)} \tag{5·16}$$

There is no difficulty now in proving by iteration from (5·10) and (5·16) that the constant has the form indicated in (5·1). Note that, in (5·9), f^{I} accounts for $(n-1)/n$ and f^{II} for $1/n$ of the total frequency.

6. Samples of 4

From (2·14), (2·15) and (3·7) the frequency for $n = 4$ is found to be

$$f_4(t_4) = \frac{\sqrt{3}}{2\pi}\int_D dx\, y^{-\frac{1}{2}}, \tag{6·1}$$

where
$$y = 2(1-x^2)^3 - (v - 3x + 5x^3)^2 \quad \text{with} \quad v = \sqrt{3}\, t_4. \tag{6·2}$$

D is the range of values of x which give non-negative values for y with $|x| \leqslant 1$. Now

$$y = -(3x^2-1)^2(3x^2-2) - 2v(5x^3-3x) - v^2, \tag{6·3}$$

from which it appears that when v is small y has two real roots near $-1/\sqrt{3}$, two imaginary roots near* $+1/\sqrt{3}$, and single roots near $+\sqrt{2}/\sqrt{3}$ and $-\sqrt{2}/\sqrt{3}$ accounting thus for all six roots. With $O(v^{\frac{3}{2}}) = 0$ the four real roots are

$$-\frac{1}{\sqrt{3}} \pm \alpha v^{\frac{1}{2}} \quad \text{with} \quad \alpha^2 = \frac{2}{9\sqrt{3}}$$
$$\pm\sqrt{\frac{2}{3}} - \frac{v}{9}.$$

Hence the integral at (6·1) may be written as the sum of five integrals

$$\int_{-\sqrt{\frac{2}{3}}-v/9}^{-\sqrt{\frac{2}{3}}+v/9} + \int_{-\sqrt{\frac{2}{3}}+v/9}^{-1/\sqrt{3}-\alpha v^{\frac{1}{2}}} + \int_{-1/\sqrt{3}+\alpha v^{\frac{1}{2}}}^{1/\sqrt{3}-\alpha v^{\frac{1}{2}}} + \int_{1/\sqrt{3}-\alpha v^{\frac{1}{2}}}^{1/\sqrt{3}+\alpha v^{\frac{1}{2}}} + \int_{1/\sqrt{3}+\alpha v^{\frac{1}{2}}}^{\sqrt{\frac{2}{3}}-v/9}.$$

Fig. 2 illustrates the division of the region of integration D. There are five divisions, numbered I–V, in what follows, in which we regard as 'principal terms' only those in $\log v$ and the constant term. Terms in $v^{\frac{1}{2}}$ will ultimately be ignored:

$$\mathrm{I} = \int_{-\sqrt{\frac{2}{3}}-v/9}^{-\sqrt{\frac{2}{3}}+v/9} = O(v^{\frac{1}{2}}),$$

$$\mathrm{II} = \int_{-\sqrt{\frac{2}{3}}+v/9}^{-1/\sqrt{3}-\alpha v^{\frac{1}{2}}} \simeq \int dx(-1+3x^2)^{-1}(2-3x^2)^{-\frac{1}{2}}$$
$$\simeq \frac{1}{4\sqrt{3}}\left\{-4\left(\frac{\sqrt{6}}{9}v\right)^{\frac{1}{2}} - 3\alpha^2 v - \log 3\alpha^2 v\right\} \simeq \int_{1/\sqrt{3}+\alpha v^{\frac{1}{2}}}^{\sqrt{\frac{2}{3}}-v/9} \simeq \mathrm{V},$$

$$\mathrm{III} = \int_{-1/\sqrt{3}-\alpha v^{\frac{1}{2}}}^{+1/\sqrt{3}-\alpha v^{\frac{1}{2}}} \simeq \int dx(1-3x^2)^{-1}(2-3x^2)^{-\frac{1}{2}} \simeq -\frac{1}{2\sqrt{3}}\log\left(\frac{3\alpha^2 v}{1-3\alpha^2 v}\right),$$

$$\mathrm{IV} = \int_{1/\sqrt{3}-\alpha v^{\frac{1}{2}}}^{1/\sqrt{3}+\alpha v^{\frac{1}{2}}} \simeq 12^{-\frac{1}{2}}\int_{-1}^{+1}(x'^2+1)^{-\frac{1}{2}}\,dx' = \frac{1}{\sqrt{3}}\sinh^{-1}1.$$

Neglecting $O(v^{\frac{1}{2}})$ we accordingly have

$$\mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}+\mathrm{V} \simeq -\frac{1}{\sqrt{3}}\log 3\alpha^2 v + \frac{1}{\sqrt{3}}\sinh^{-1}1 + 2C. \tag{6·4}$$

The constant $2C$ derives from additional terms in integrals II, III and V:

$$\mathrm{III} = \int_{-\gamma}^{+\gamma} dx\{(1-3x^2)^2(2-3x^2) - \overbrace{2v(5x^3-3x)}^{z} - v^2\}^{-\frac{1}{2}}$$

with
$$\gamma = 1/\sqrt{3} - \alpha v^{\frac{1}{2}}.$$

* In a sense which will be obvious from Fig. 2.

We have already taken account of the term in III found when v is zero. The constant C derives from the even powers of z in the formal expansion of the denominator of the integral element—the odd powers vanish by symmetry. Setting

$$x = \frac{1}{\sqrt{3}} - x', \quad 1-3x^2 \simeq 2\sqrt{3}\,x', \quad 2-3x^2 \simeq 1, \quad 5x^3-3x \simeq -\tfrac{4}{3}\sqrt{3}.$$

On expansion of III,

$$C \simeq 2\sum_{k=1}^{\infty}\int_{\alpha v^{\frac{1}{2}}}^{1/\sqrt{3}} dx'\,x'^{-4k-1}(2v)^{2k}(2\sqrt{3})^{-4k-1}C_{2k},$$

where C_{2k} are the even-order coefficients in the expansion of $(1+z)^{-\frac{1}{2}}$, i.e. C_{2k} is the coefficient of z^{2k}. On integration we are interested only in the value at the lower limit $\alpha v^{\frac{1}{2}}$, for all terms at the upper limit (and certain terms at the lower limit) are $O(v^{\frac{1}{2}})$ at least. Hence

$$C = \frac{1}{4\sqrt{3}}\sum_{k=1}^{\infty} C_{2k}/k.$$

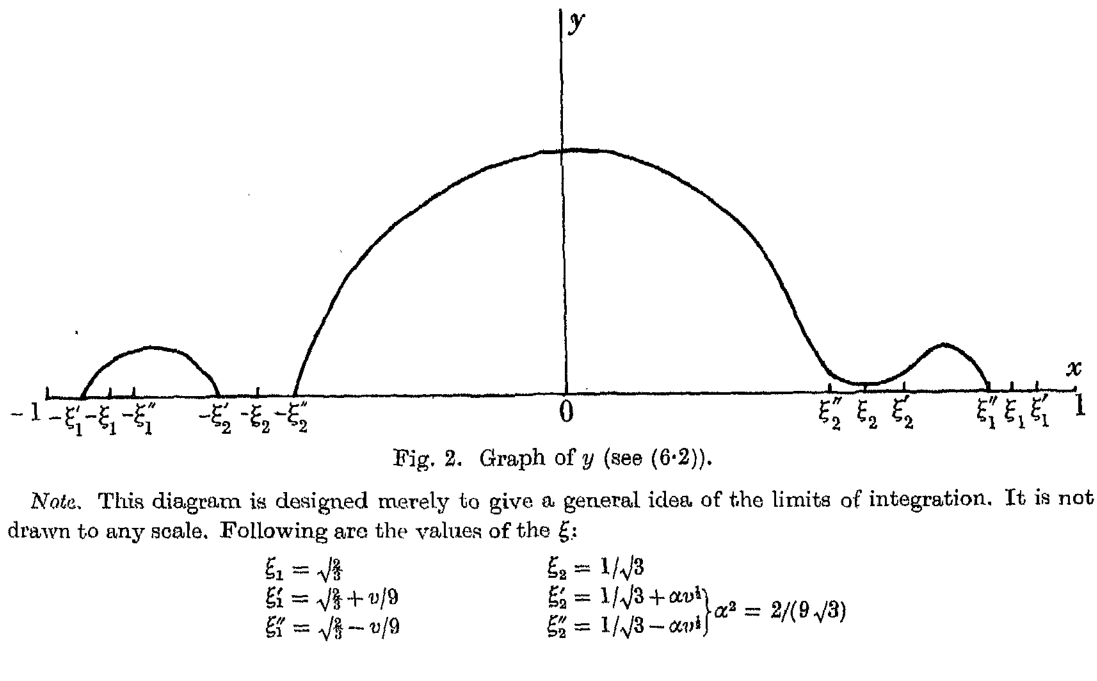


Fig. 2. Graph of y (see (6·2)).

Note. This diagram is designed merely to give a general idea of the limits of integration. It is not drawn to any scale. Following are the values of the ξ:

$$\begin{array}{ll} \xi_1 = \sqrt{\tfrac{2}{3}} & \xi_2 = 1/\sqrt{3} \\ \xi_1' = \sqrt{\tfrac{2}{3}} + v/9 & \left.\begin{array}{l}\xi_2' = 1/\sqrt{3} + \alpha v^{\frac{1}{2}} \\ \xi_2'' = 1/\sqrt{3} - \alpha v^{\frac{1}{2}}\end{array}\right\}\alpha^2 = 2/(9\sqrt{3}) \\ \xi_1'' = \sqrt{\tfrac{2}{3}} - v/9 & \end{array}$$

Similarly II + V also yield C giving a constant additional term of $2C$.

Now

$$\begin{aligned} C &= \frac{1}{\sqrt{3}}\sum_{k=1}^{\infty}\frac{C_{2k}}{4k} = \frac{1}{\sqrt{3}}\int_0^1 \frac{dx}{x}(C_2x^4 + C_4x^8 + \ldots) \\ &= \frac{1}{\sqrt{3}}\int_0^1 -\frac{dx}{x} + \frac{1}{2\sqrt{3}}\int_0^1 \frac{dx}{x}\{(1+x^2)^{-\frac{1}{2}} + (1-x^2)^{-\frac{1}{2}}\} \\ &= \frac{1}{\sqrt{3}}\left[-\log x + \tfrac{1}{4}\log\frac{(1+x^2)^{\frac{1}{2}}-1}{(1+x^2)^{\frac{1}{2}}+1} + \tfrac{1}{4}\log\frac{1-(1-x^2)^{\frac{1}{2}}}{1+(1-x^2)^{\frac{1}{2}}}\right]_{x=0}^{x=1} \\ &= \frac{1}{\sqrt{3}}\{\log 2 + \tfrac{1}{2}\log(2^{\frac{1}{2}}-1)\}. \end{aligned} \qquad (6{\cdot}5)$$

All logs are to base e (unless otherwise indicated in what follows). Hence

$$f_4(t_4) \simeq 0{\cdot}372646 - \frac{1}{2\pi}\log v \qquad (6{\cdot}6)$$

$$= 0{\cdot}285222 - 0{\cdot}366466 \log_{10} |\, t_4 \,|, \qquad (6{\cdot}7)$$

since $$v = \sqrt{3}\, t_4.$$

A. T. McKay (1933), from a different approach, gave the log term in (6·7), and, as a rough approximation to the constant term, the value 0·311568. He also showed that an expression of the form (6·7) accounted for most of the frequency, a fact of great importance. Assume that the residual term is of form

$$A\,|\, t_4 \,|^{\frac{1}{2}} + B\,|\, t_4 \,|,$$

and find A and B from

(i) $f_4(2/\sqrt{3}) = \frac{1}{2}\sqrt{3}$ (from (5·1); also McKay (1933)),

(ii) total frequency is unity,

giving

$$f_4(t_4) = 0{\cdot}285222 - 0{\cdot}366466 \log_{10} |\, t_4 \,| - 0{\cdot}009178\,|\, t_4 \,|^{\frac{1}{2}} + 0{\cdot}031359\,|\, t_4 \,|. \qquad (6{\cdot}8)$$

For algebraic manipulation at the next stage the form of residual $A'\,|\, t_4 \,| + B' t_4^2$ will be found more convenient, however, with A' and B' also determined from (i) and (ii). In this form

$$f_4(t_4) = 0{\cdot}285222 - 0{\cdot}159155 \log |\, t_4 \,| + 0{\cdot}014275\,|\, t_4 \,| + 0{\cdot}007398 t_4^2. \qquad (6{\cdot}9)$$

Note the smallness of the coefficients A, B, A' and B' in (6·8) and (6·9).

In the following table the first four even moments as derived from frequencies (6·8) and (6·9) are compared with the actual values as derived from the formulae (1·2). Both formulae yield excellent approximations, with (6·8) always superior to (6·9) however. Either formula can obviously be used with complete confidence for deriving the probability points. The frequency graph in Fig. 4 is derived from (6·8) which should also be used for the computation of the probability points.

Moment	Actual	Formula (6·8)	Formula (6·9)
μ_2	0·342857	0·342930	0·342470
μ_4	0·258941	0·258979	0·258606
μ_6	0·240503	0·240263	0·240205
μ_8	0·245940	0·245949	0·246662

7. Samples of 5

After many computational experiments the method used for determining the frequency $f_5(t_5)$ was as follows:

(1) Using (2·14) with form (6·9) for $f_4(t_4)$, central and link ordinates, i.e. $f_5(0)$ and $f_5(1/\sqrt{6})$ were computed.

(2) The approximate value of $f_5(t_5)$ near $t_5 = 1/\sqrt{6} + 0$ was found in the form

$$f_5(1/\sqrt{6}) + M(t_5 - 1/\sqrt{6})^{\frac{1}{2}},$$

M being known.

(3) The two zonal curves were found (i) passing through $(0, f_5(0))$ and $(1/\sqrt{6}, f_5(1/\sqrt{6}))$ with $f_5'(0) = 0$ and (ii) passing through $(1/\sqrt{6}, f_5(1/\sqrt{6}))$ and with the required form at $1/\sqrt{6}+0$ (i.e. as at (2) above) and at the limit $(\frac{3}{2}-0)$ so that $\mu_0\ (=1), \mu_2, \mu_4$ and μ_6 have the exact values as given for $n = 5$ by the formulae (1·2).

Setting then

$$f_4(t_4) = 0{\cdot}285222 - 0{\cdot}159155 \log_e |t_4| + R(t_4), \tag{7·1}$$

with

$$t_4 = \frac{6}{\sqrt{5}} \frac{(x^3 - \frac{1}{2}x + \frac{1}{3}t_5)}{(1-x^2)^{\frac{3}{2}}}, \tag{7·2}$$

and

$$R(t_4) = 0{\cdot}014275\,|t_4| + 0{\cdot}007398 t_4^2, \tag{7·3}$$

we have

$$f_5(t_5) = \frac{4}{\pi\sqrt{5}} \int_\lambda^\mu \frac{dx}{1-x^2} f_4(t_4), \tag{7·4}$$

the limits of integration being λ (negative) and μ (positive) which are the values of x, from (7·2), corresponding respectively to $t_4 = -\frac{1}{2}\sqrt{3}$ and $t_4 = +\frac{1}{2}\sqrt{3}$. We shall be concerned only with the case $t_5 \leqslant 1/\sqrt{6}$ when $t_4(x)$ has three real roots β, α and γ of which β is negative and α and γ are positive. For (7·4) the following are required

$$\int_\lambda^\mu \frac{dx}{1-x^2} = \tfrac{1}{2}\log\left(\frac{1+\mu}{1+\lambda}\frac{1-\lambda}{1-\mu}\right), \tag{7·5}$$

$$\begin{aligned}\int_\lambda^\mu \frac{dx \log(1-x^2)}{(1-x^2)} = {} & \tfrac{1}{4}\{\log^2(1+\mu) - \log^2(1+\lambda) - \log^2(1-\mu) + \log^2(1-\lambda)\} \\ & + \tfrac{1}{2}\{\log(1+\mu)\log(1-\mu) - \log(1+\lambda)\log(1-\lambda)\} \\ & + \log 2 \log\left(\frac{1-\lambda}{1-\mu}\right) + J\left(\frac{1+\mu}{2}\right) - J\left(\frac{1+\lambda}{2}\right),\end{aligned} \tag{7·6}$$

$$\begin{aligned}\int_\lambda^\mu \frac{dx}{1-x^2}\log\left|x^3 - \frac{x}{2} + \frac{t_5}{3}\right| = {} & \tfrac{1}{2}\log\left(\frac{1-\lambda}{1-\mu}\right)\log(1-\alpha)(1-\beta)(1-\gamma) \\ & - \tfrac{1}{2}\log\left(\frac{1+\lambda}{1+\mu}\right)\log(1+\alpha)(1+\beta)(1+\gamma) + \frac{1}{2}\left\{\sum_{i=1}^{6} I(\kappa_i) + \sum_{j=7}^{12} J(\kappa_j)\right\},\end{aligned} \tag{7·7}$$

with

$$\begin{array}{lll} \kappa_1 = (\alpha-\lambda)/(1-\alpha), & \kappa_5 = (\beta-\lambda)/(1-\beta), & \kappa_9 = (\gamma-\lambda)/(1+\gamma), \\ \kappa_2 = (\mu-\alpha)/(1+\alpha), & \kappa_6 = (\mu-\beta)/(1+\beta), & \kappa_{10} = (\mu-\gamma)/(1-\gamma), \\ \kappa_3 = (\gamma-\lambda)/(1-\gamma), & \kappa_7 = (\alpha-\lambda)/(1+\alpha), & \kappa_{11} = (\beta-\lambda)/(1+\beta), \\ \kappa_4 = (\mu-\gamma)/(1+\gamma), & \kappa_8 = (\mu-\alpha)/(1-\alpha), & \kappa_{12} = (\mu-\beta)/(1-\beta), \end{array}$$

and

$$I(\kappa) = \int_0^\kappa \frac{dx \log x}{1+x} = \log\kappa \log(1+\kappa) - \psi(\kappa),$$

$$J(\kappa) = \int_0^\kappa \frac{dx \log x}{1-x} = -\log\kappa\log(1-\kappa) - \phi(\kappa),$$

$$\left.\begin{aligned}\phi(\kappa) &= \frac{\kappa}{1^2} + \frac{\kappa^2}{2^2} + \frac{\kappa^3}{3^2} + \dots \\ \psi(\kappa) &= \frac{\kappa}{1^2} - \frac{\kappa^2}{2^2} + \frac{\kappa^3}{3^2} - \dots\end{aligned}\right\} \text{ when } \kappa \leqslant 1.$$

It is useful to note that $\phi(1) = 1{\cdot}644934 = 2\psi(1)$. The functions $\phi(\kappa)$ and $\psi(\kappa)$ do not appear to be tabulated. By fitting curves to their values for equally spaced intervals of 0·05 from 0 to 0·5 the following very close approximations are found, applicable for $x \leqslant \frac{1}{2}$:

$$\phi(x) = 1{\cdot}000567x + 0{\cdot}233454x^2 + 0{\cdot}186052x^3,$$
$$\psi(x) = 0{\cdot}999835x - 0{\cdot}244220x^2 + 0{\cdot}077024x^3.$$

When $1 \geqslant \kappa > \frac{1}{2}$ the following formulae can be used:

$$\phi(\kappa) = \phi(1) - \log\kappa \log(1-\kappa) - \phi(1-\kappa),$$
$$\psi(\kappa) = \tfrac{1}{2}\phi(1) + \log\kappa \log(1+\kappa) - \phi(1-\kappa) + \tfrac{1}{2}(1-\kappa^2).$$

When $\kappa > 1$ we use

$$\phi(\kappa) = 2\phi(1) - \tfrac{1}{2}\log^2 1/\kappa - \phi(1/\kappa),$$
$$\psi(\kappa) = 2\psi(1) + \tfrac{1}{2}\log^2 1/\kappa - \psi(1/\kappa).$$

Another useful formula is

$$\phi(\kappa) = \psi(\kappa) + \tfrac{1}{2}\phi(\kappa^2).$$

The algebra of the contribution to (7·4) from $R(t_4)$ is without mathematical interest. From the formula the following were the values found for the central frequency and the second link frequency:

$$f_5(0) = 0{\cdot}606563;\quad f_5(1/\sqrt{6}) = 0{\cdot}599069. \tag{7·8}$$

The moments, computed by (2·1), with $n = 5$, are

$$\mu_0 = 1,\quad \mu_2 = 0{\cdot}375,\quad \mu_4 = 0{\cdot}361607,\quad \mu_6 = 0{\cdot}474609,\quad \mu_8 = 0{\cdot}719382. \tag{7·9}$$

Computation by approximate integration of certain of the ordinates gave evidence of marked irregularity near the link $t_5 = 1/\sqrt{6}$. In consequence, it seemed desirable to try to find a term (in addition to the constant given at (7·8)) of the expansion of $f_5(t_5)$ near $1/\sqrt{6}+0$. Setting

$$t_5 = \frac{1}{\sqrt{6}} + t, \tag{7·10}$$

where t is small and positive—we shall be interested only in a term in $t^{\frac{1}{2}}$—we find

$$f_5(t_5) \simeq \frac{4}{\pi\sqrt{5}}\left\{\int_{\lambda+\lambda' t}^{v-At^{\frac{1}{2}}} + \int_{v+At^{\frac{1}{2}}}^{\mu+\mu' t}\right\}\frac{dx}{1-x^2} f_4(t_4). \tag{7·11}$$

The values $v \pm At^{\frac{1}{2}}$ are the abscissae of the points at which the curve $t_4 = t_4(x)$, given by (7·2), intersects the t_4 link line $t_4 = 2/\sqrt{3}$ near $x = v = -\frac{1}{4}\sqrt{6}$. It can easily be shown that

$$A^2 = \frac{\sqrt{6}}{12}. \tag{7·12}$$

We are not concerned with the values $(\lambda + \lambda' t)$ and $(\mu + \mu' t)$ which are the abscissae corresponding to the intersection of $t_4 = t_4(x)$ with $t_4 = -2/\sqrt{3}$ and its third intersection with $t_4 = 2/\sqrt{3}$. Remembering that at the latter link $f_4(t_4)$ has the value $1/(2\sqrt{3})$, the *integral-free* term in $t^{-\frac{1}{2}}$ in the first derivative $f_5'(t_5)$ of $f_5(t_5)$ is

$$\frac{4}{\pi\sqrt{5}}\frac{1}{1-v^2}\{f_4(v-At^{\frac{1}{2}}) \times -\tfrac{1}{2}At^{-\frac{1}{2}} + f_4(v+At^{\frac{1}{2}}) \times -\tfrac{1}{2}At^{-\frac{1}{2}}\} = -\frac{16A\sqrt{15}}{75\pi}t^{-\frac{1}{2}}. \tag{7·13}$$

Also we have to consider the *integral* term in $f_5'(t_5)$. For this purpose, from (7·10) and (7·2),

$$t_4 = \frac{6}{\sqrt{5}}\left\{(x+\sqrt{\tfrac{2}{3}})\left(x-\frac{1}{\sqrt{6}}\right)^2 + \frac{t}{3}\right\}(1-x^2)^{-\frac{3}{2}}$$
$$\simeq \frac{6}{\sqrt{5}}\left(x+\sqrt{\tfrac{2}{3}}+\frac{2t}{9}\right)\left\{\left(x-\frac{1}{\sqrt{6}}-\frac{t}{9}\right)^2 + \frac{\sqrt{6}}{9}t\right\}(1-x^2)^{-\frac{3}{2}}.$$

Remembering (7·1), it can be shown that the only term in (7·11) from which a term in $t^{\frac{1}{2}}$ can come is approximately

$$-\frac{4}{\pi\sqrt{5}}\frac{1}{2\pi}\int_{\lambda}^{\mu}\frac{dx}{1-x^2}\log\left\{\left(x-\frac{1}{\sqrt{6}}-\frac{t}{9}\right)^2+\frac{\sqrt{6}}{9}t\right\}, \tag{7·14}$$

λ and μ, the limiting values, being respectively negative and positive.

Differentiating (7·14) in respect of t we find

$$-\frac{2}{\pi^2\sqrt{5}}\int_{\lambda}^{\mu}\frac{dx}{1-x^2}\left\{\left(x-\frac{1}{\sqrt{6}}-\frac{t}{9}\right)^2+\frac{\sqrt{6}}{9}t\right\}^{-1}\left\{\frac{\sqrt{6}}{9}-\frac{2}{9}\left(x-\frac{1}{\sqrt{6}}-\frac{t}{9}\right)\right\}. \tag{7·15}$$

Changing variables by $$x-\left(\frac{1}{\sqrt{6}}+\frac{t}{9}\right) = \frac{6^{\frac{1}{4}}}{3}t^{\frac{1}{2}}y$$

and letting t tend towards $+0$, we find for the term in $t^{-\frac{1}{2}}$

$$-\frac{2}{\pi^2\sqrt{5}}\left(\frac{\pi 6^{\frac{1}{4}}2}{5}t^{-\frac{1}{2}}\right). \tag{7·16}$$

Adding (7·13) and (7·16) we find

$$-\frac{1}{\pi}5^{-\frac{1}{2}}2^{\frac{9}{4}}3^{-\frac{3}{4}}t^{-\frac{1}{2}}.$$

On integrating we find for the term in $t^{\frac{1}{2}} = \left(t_5-\frac{1}{\sqrt{6}}\right)^{\frac{1}{2}}$

$$-\frac{1}{\pi}5^{-\frac{1}{2}}2^{\frac{13}{4}}3^{-\frac{3}{4}}\left(t_5-\frac{1}{\sqrt{6}}\right)^{\frac{1}{2}} = -0{\cdot}594117\left(t_5-\frac{1}{\sqrt{6}}\right)^{\frac{1}{2}}. \tag{7·17}$$

From (5·1) the value of $f_5(x)$ near $x = \frac{3}{2}$ is

$$0{\cdot}219166(\tfrac{3}{2}-x)^{\frac{1}{2}}, \tag{7·18}$$

where x is usually written for simplicity instead of t_5 in the remainder of this section.

Having regard to (7·8) and to the fact that, from § 4, $f_5'(0) = 0$, in the half-zone $(0-1/\sqrt{6})$, $f_5(x) = F(x)$ must be of form

$$F(x) = 0{\cdot}606563+a_2x^2+a_3x^3+a_4x^4. \tag{7·19}$$

The first relation between the coefficients is found by giving expression to the fact that $y = F(x)$ passes through the link-point $(1/\sqrt{6}, 0{\cdot}599069)$:

$$0{\cdot}166667a_2+0{\cdot}068041a_3+0{\cdot}027778a_4 = -0{\cdot}007494. \tag{7·20}$$

In the zone $(1/\sqrt{6}-\frac{3}{2})$ assume that

$$f_5(x) = G(x) = -0{\cdot}594117\left(x-\frac{1}{\sqrt{6}}\right)^{\frac{1}{2}}+0{\cdot}219166(\tfrac{3}{2}-x)^{\frac{1}{2}}+b_0+b_1(\tfrac{3}{2}-x)$$
$$+b_2(\tfrac{3}{2}-x)^2+b_3(\tfrac{3}{2}-x)^3, \tag{7·21}$$

designed to conform with requirements (7·17) and (7·18). Since $y \doteq G(x)$ must pass through $(\frac{3}{2}, 0)$,

$$b_0 = 0{\cdot}594117\left(\frac{3}{2}-\frac{1}{\sqrt{6}}\right)^{\frac{1}{2}} = 0{\cdot}620775. \tag{7·22}$$

Taking the value of b_0 into account and giving algebraic expression to $y = G(x)$ passing through $(1/\sqrt{6}, 0{\cdot}599069)$, we find

$$1{\cdot}091752b_1+1{\cdot}191922b_2+1{\cdot}301283b_3 = -0{\cdot}250706. \tag{7·23}$$

To find the six coefficients a_2, a_3, a_4 (in (7·19)) and b_1, b_2, b_3 (in (7·21)), we have, so far, found two equations, (7·20) and (7·23). The remaining four equations are found by equating the total frequency to unity and the first three even moments to their true values given at (7·9), i.e. setting

$$\tfrac{1}{2}\mu_{2k} = \int_0^{1/\sqrt{6}} dx\, x^{2k} F(x) + \int_{1/\sqrt{6}}^{\frac{3}{2}} dx\, x^{2k} G(x) \quad (k = 0, 1, 2, 3).$$

On substituting for a_3 given by (7·20), for b_0 given by (7·22) and for b_3 given by (7·23), we find the four equations in a_2, a_3, b_1, b_2:

$$\left.\begin{aligned}
0{\cdot}0^2 90721 a_2 + 0{\cdot}0^2 139999 a_3 + 0{\cdot}297981 b_1 + 0{\cdot}108441 b_2 &= -0{\cdot}071173,\\
0{\cdot}0^3 64802 a_2 + 0{\cdot}0^3 110237 a_3 + 0{\cdot}268332 b_1 + 0{\cdot}082590 b_2 &= -0{\cdot}066293,\\
0{\cdot}0^4 5901 a_2 \;\; + 0{\cdot}0^4 107175 a_3 + 0{\cdot}303248 b_1 + 0{\cdot}079086 b_2 &= -0{\cdot}076019,\\
0{\cdot}0^5 6368 a_2 \;\; + 0{\cdot}0^5 1169 a_3 \quad + 0{\cdot}400090 b_1 + 0{\cdot}089674 b_2 &= -0{\cdot}101300.
\end{aligned}\right\} \qquad (7{\cdot}24)$$

On solution (and checking by substitution) the coefficients are found to give finally the following frequencies:

$$\left.\begin{aligned}
&\text{Zone} \qquad\qquad\qquad\qquad f_5(x)\\
&0 - 1/\sqrt{6}: 0{\cdot}606563 - 0{\cdot}3307x^2 + 3{\cdot}1955x^3 - 6{\cdot}1129x^4,\\
&1/\sqrt{6} - \tfrac{3}{2}: 0{\cdot}620775 - 0{\cdot}594117\left(x - \frac{1}{\sqrt{6}}\right)^{\frac{1}{2}} + 0{\cdot}219166(\tfrac{3}{2} - x)^{\frac{1}{2}} - 0{\cdot}268273(\tfrac{3}{2} - x)\\
&\qquad\qquad + 0{\cdot}067263(\tfrac{3}{2} - x)^2 - 0{\cdot}029195(\tfrac{3}{2} - x)^3,
\end{aligned}\right\} \qquad (7{\cdot}25)$$

with $x = t_5$.

The extremely interesting form of the frequency curve may be observed from Fig. 5. In the first half-zone the frequency shows but little variation: the curve declines to a minimum of 0·6058 at $x = 0{\cdot}0894$ then rises to a maximum of 0·6136 at $x = 0{\cdot}3027$. It then recedes to the link $1/\sqrt{6}$, where it assumes the value 0·5991. As one type of check on the reliability of the results in general, some ordinates were computed directly (i.e. using (7·4) and (7·1)), or by approximate integration using (7·4) and (6·8) and compared with the ordinates computed from (7·25) to the following effect:

Trial value of t_5	Value of frequency	
	By approx. integration	By (7·25)
0·15	0·6069	0·6068
0·3	0·6106	0·6136
0·6	0·3650	0·3603
0·9	0·2232	0·2308
1·2	0·1377	0·1371

Except perhaps for the frequency at $t_5 = 0{\cdot}9$, the correspondence is satisfactory; there can be little doubt that the more accurate figures are those from (7·25).

As a stringent test of the accuracy of the frequency the 8th moment μ_8' was computed from the empirical curves at (7·25) and compared with the actual value given at (7·9):

$$\mu_8' = 0{\cdot}7191, \quad \mu_8 = 0{\cdot}7194.$$

Even as the figures stand the check is decisive: it should be added that the 4th place of decimals in μ_8' is suspect to the approximation used.

8. SAMPLES OF 6

In this case the links are 0, $1/\sqrt{2}$ and $4/\sqrt{5}$, and the link frequencies at the first two were found by approximate integration using form (2·13) with t_6 given by (2·8). For this purpose, drawings were made of the two sections of $f_5(t_5)$ on a scale sufficient to ensure that an ordinate read for any abscissa would be correct probably to the 3rd place of decimals. For intervals of 1°, values of t_5 were computed over the whole range by (2·8) (for t_6 given), and graphically the value of $f_5(t_5)$ was read off for each t_5. Hundreds of readings had to be made, but actually the work, with a little practice, was rapid and accurate, the entries being practically self-checking. The Gregory formula (using 2 correction terms) was used to give the following results:

$$f_6(0) = 0\cdot6889; \quad f_6(1/\sqrt{2}) = 0\cdot3247. \tag{8·1}$$

The two zonal frequency curves, say $y = F(x)$ in $(0 - 1/\sqrt{2})$ and $y = G(x)$ in $(1/\sqrt{2} - 4/\sqrt{5})$, *writing x instead of t_6* must have the following properties:

$$\left.\begin{array}{ll}
\text{(i)} & F(0) = 0\cdot6889, \\
\text{(ii)} & F'(0) = 0 \ (\S\,4), \\
\text{(iii)} & F(1/\sqrt{2}) = G(1/\sqrt{2}) = 0\cdot3247, \\
\text{(iv)} & F'(1/\sqrt{2}) = G'(1/\sqrt{2}) \ (\S\,3), \\
\text{(v)} & G(x) \simeq 5(4/\sqrt{5} - x)/36 \ (\text{from } (5\cdot1)).
\end{array}\right\} \tag{8·2}$$

The curves were

$$\left.\begin{array}{l}
F(x) = 0\cdot6889 + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5, \\
G(x) = \frac{5}{36}(\beta - x) + b_2(\beta - x)^2 + b_3(\beta - x)^3 + b_4(\beta - x)^4 \text{ with } \beta = 4/\sqrt{5}.
\end{array}\right\} \tag{8·3}$$

The exact moments are

$$\mu_0 = 1, \quad \mu_2 = 0\cdot380952, \quad \mu_4 = 0\cdot409191, \quad \mu_6 = 0\cdot642924, \quad \mu_8 = 1\cdot219892. \tag{8·4}$$

It is proposed to compute the seven coefficients in (8·3) using (8·2) and (8·4). Now, with a curve of the type of $f_6(x)$, where much of the frequency is at the ends it is evident that the contribution from the zone $(0 - 1/\sqrt{2})$ to the higher moments μ_6 and μ_8 is exceedingly minute: this property was utilized to divide the single series of seven equations into two series of three and four equations using the following device: Approximate $F(x)$ by a curve $F_1(x)$ given by

$$F_1(x) = 0\cdot6889 + a_2'x^2, \tag{8·5}$$

finding a_2' simply by passing $y = F_1(x)$ through $(1/\sqrt{2}, 0\cdot3247)$ giving $a_2' = -0\cdot7284$. If μ_{2s} is the moment, let μ_{2s}' and μ_{2s}'' be the contributions from $F(x)$ and $G(x)$ respectively so that $\mu_{2s} = \mu_{2s}' + \mu_{2s}''$. Let ν_{2s}' be the estimate, using $F_1(x)$, of μ_{2s}'. For $s = 2, 3, 4$ the values are

$$\nu_4' = 0\cdot030318, \quad \nu_6' = 0\cdot009360, \quad \nu_8' = 0\cdot003397,$$

which, subtracted from the corresponding μ_{2s} given by (8·4), give very close estimates of μ_{2s}'', which involve only b_2, b_3, b_4. The equations in order

$$\left.\begin{array}{ll}
(1) & 1\cdot170178b_2 + 1\cdot265837b_3 + 1\cdot369316b_4 = 0\cdot174500, \\
(2) & 0\cdot726980b_2 + 0\cdot378949b_3 + 0\cdot235502b_4 = 0\cdot058771, \\
(3) & 1\cdot260067b_2 + 0\cdot534702b_3 + 0\cdot289663b_4 = 0\cdot091217,
\end{array}\right\} \tag{8·6}$$

are found from (iii) at (8·2), from μ_6'' and from μ_8''.

The equations in the a's are

$$\left.\begin{array}{llll} (4) & 0{\cdot}5a_2 + 0{\cdot}353553a_3 + 0{\cdot}25a_4 + 0{\cdot}176777a_5 & = & -0{\cdot}3642, \\ (5) & 1{\cdot}414214a_2 + 1{\cdot}5a_3 + 1{\cdot}414214a_4 + 1{\cdot}25a_5 & = & -0{\cdot}653179, \\ (6) & 0{\cdot}117851a_2 + 0{\cdot}0625a_3 + 0{\cdot}035355a_4 + 0{\cdot}020833a_5 & = & -0{\cdot}115790, \\ (7) & 0{\cdot}035355a_2 + 0{\cdot}020833a_3 + 0{\cdot}012627a_4 + 0{\cdot}007813a_5 & = & -0{\cdot}031433, \end{array}\right\} \quad (8{\cdot}7)$$

where (4) is from (iii) at (8·2), (5) from (iv) at (8·2), (6) from the total frequency $= \frac{1}{2}$ and (7) from variance $= \mu_2$.

The solutions of (8·6) and (8·7) yield the following frequencies:

$$\left.\begin{array}{ll} \text{Zone} & f_6(x) \\ 0 - 1/\sqrt{2} & : 0{\cdot}6889 - 1{\cdot}2715x^2 - 2{\cdot}6073x^3 + 9{\cdot}5669x^4 - 6{\cdot}7790x^5, \\ 1/\sqrt{2} - 4/\sqrt{5} & : \frac{5}{36}(\beta - x) + 0{\cdot}047068(\beta - x)^2 + 0{\cdot}024897(\beta - x)^3 + 0{\cdot}064198(\beta - x)^4, \\ & \text{with } \beta = 4/\sqrt{5}, \end{array}\right\} \quad (8{\cdot}8)$$

with $x = t_6$.

As a check, the 4th moment computed from the foregoing curves gave 0·4108 as compared with the actual $\mu_4 = 0{\cdot}4092$, an error of 0·38 %. This is not of any importance from the viewpoint of the computation of the probability points, but it illustrates how, using the integral iteration method as generally in this paper, the momental check reveals increasing discrepancies with increasing n.

It might be thought that by constructing empirically 'almost any' symmetrical frequency curve, so that say the 0th, 2nd and 4th moments have the true values, we shall ensure that the subsequent even moments computed from such an empirical curve will approximate closely to the corresponding true values. That this is not the case may be seen by computing the 6th and 8th moments by the well-known Karl Pearson iteration formula,* where μ_2 and μ_4 have their true values, for $\sqrt{b_1}$ with $n = 6$:

	Actual	Karl Pearson iteration	Percentage discrepancy
$\beta_4 = \mu_6/\mu_2^3$	11·6291	12·4984	+10·7
$\beta_6 = \mu_8/\mu_2^4$	57·9214	73·3990	+26·7

Even when, in the Pearson iteration for β_6, one gives β_4 the correct value, we find a percentage discrepancy of 17·9. These percentages place in perspective the minuteness of the percentage errors found in using the higher momental check as it is used throughout this paper.

It is an interesting question of general import whether in work of this kind the arduous and potentially erroneous computation (by integral iteration) of the central and link frequencies could be dispensed with, and reliance placed entirely on the moments, together with the functional properties of the frequencies, which, of course, merely represent an elaboration of the Karl Pearson approach. In this connexion a couple of experiments were made on the $\sqrt{b_1}$ frequency for $n = 6$.

* *Tables for Statisticians and Biometricians*, Part I, 2nd ed., p. xi.

For the first experiment, the two zonal curves were assumed to have the correct order of contact, the correct form, (8·2) (v), at the limit of range and the correct values of μ_0 $(=1)$, μ_2, μ_4 and μ_6. The equations are

$$\begin{aligned}
&\text{Zone} \\
&0-1/\sqrt{2} \quad : F_1(x) = 0\cdot659844 - 1\cdot075618x^2 + 0\cdot555991x^3, \\
&1/\sqrt{2}-4/\sqrt{5} : G_1(x) = \tfrac{5}{36}(\beta-x) + 0\cdot080560(\beta-x)^2 - 0\cdot085469(\beta-x)^3 \\
&\qquad\qquad + 0\cdot133119(\beta-x)^4 \text{ with } \beta = 4/\sqrt{5}.
\end{aligned} \qquad (8\cdot9)$$

This gives a central frequency 0·6598 compared with the computed frequency (by (8·1)) of 0·6889. In all the circumstances the difference is not important. The 8th moment, μ_8'', from (8·9), is 1·217706, or $-0\cdot18$ % in error.

The second experiment contemplated the frequency as a single-curve system with correct first derivative $(-\frac{5}{36})$ at the limit and with correct μ_0, μ_2, μ_4, μ_6. The curve is

$$F_2(x) = 0\cdot669426 - 1\cdot51097x^2 + 1\cdot53854x^3 - 0\cdot60545x^4 + 0\cdot085157x^5, \qquad (8\cdot10)$$

which has the properties: (i) the central ordinate 0·6694 is close to the actual; (ii) limit value from curve scarcely differed from the actual since $F_2(4/\sqrt{5}) = 0\cdot0015$; (iii) μ_8''' from curve $= 1\cdot2237$, an error of 0·31 %.

All the systems (8·8), (8·9) or (8·10) yield probability points which differ very little. For instance, in the three cases, the 5 % point is given by

System	5 % probability
(8·8)	1·0432
(8·9)	1·0385
(8·10)	1·0384

The practical identity of the latter two is due to the fact that the frequencies were derived on very similar hypotheses: it does not mean that the result is more reliable than that from (8·8) which, assuming the accuracy of the calculation of the link ordinates, must be deemed to be the most correct and is adopted for the iteration to the $n = 7$ stage. Nevertheless, these experiments convey the hint of general application that if we know (i) a number of moments, (ii) the limits of range and the frequency form at the limits of range, and (iii) that the amount of frequency near the limits of range is not negligible, we will probably be in a position to estimate with fair accuracy the points of low probability. For this, however, hypothesis (iii) is essential: it has no value from the computational point of view if the frequency near the limits is negligible. This point is discussed further in § 10.

9. Samples of 7

The functional properties of the curves at the stage are as follows. Let the three links be denoted by α, β, γ, so that

$$\alpha = 1/\sqrt{12}, \quad \beta = 3/\sqrt{10}, \quad \gamma = 5/\sqrt{6}. \qquad (9\cdot1)$$

Denoting t_7 by x, set

$$y = x - \alpha, \quad z = \gamma - x, \qquad (9\cdot2)$$

and let the curves in the half-zone $(0-1/\sqrt{12})$, and in the zones $(1/\sqrt{12}-3/\sqrt{10})$ and $(3/\sqrt{10}-5/\sqrt{6})$ be denoted respectively by $F(x)$, $G(y)$ and $H(z)$. We then have

$$\left.\begin{array}{rl}
\text{(i)} & F(0) = 0{\cdot}6781 = A, \\
\text{(ii)} & F'(0) = 0, \\
\text{(iii)} & F(\alpha) = G(0) = 0{\cdot}5870 = B, \\
\text{(iv)} & F'(\alpha) = G'(0), \\
\text{(v)} & F''(\alpha) = G''(0), \\
\text{(vi)} & G(\beta-\alpha) = H(\gamma-\beta) = 0{\cdot}1838 = C, \\
\text{(vii)} & G'(\beta-\alpha) = -H'(\gamma-\beta), \\
\text{(viii)} & H(z) = Dz^{\frac{3}{2}} + c_2 z^2 + c_3 z^3 \text{ with } D = 0{\cdot}078091.
\end{array}\right\} \qquad (9{\cdot}3)$$

The central and link ordinates A, B and C at (i), (iii) and (vi), were derived by the Gregory formula from (2·14), using intervals of 0·01, 0·025 and 0·05 at different sections of the integral range. The equalities in the derivatives at the links are in accordance with order of contact requirements (§ 3). The first term on the right of (viii) is from (5·1) with $n = 7$.

Conditions (9·3) determine the form of the polynomials:

$$\left.\begin{aligned}
F(x) &= A + a_2 x^2 + a_4 x^4, \\
G(y) &= B + (2a_2\alpha + 4a_4\alpha^3)\, y + \tfrac{1}{2}(2a_2 + 12a_4\alpha^2)\, y^2 + b_3 y^3 + b_4 y^4, \\
H(z) &= Dz^{\frac{3}{2}} + c_2 z^2 + c_3 z^3,
\end{aligned}\right\} \qquad (9{\cdot}4)$$

with $x = t_7$.

$F(x)$ is taken as an even function of x because it is symmetrical in the zone $(-1/\sqrt{12}$ to $+1/\sqrt{12})$. This should have been done in the case of $n = 5$; neglect to do so was not serious enough to render recalculation necessary.

The moments used were:

$$\mu_0 = 1, \quad \mu_2 = 0{\cdot}375, \quad \mu_4 = 0{\cdot}421875, \quad \mu_6 = 0{\cdot}733487. \qquad (9{\cdot}5)$$

Using (9·3) in conjunction with μ_0 and μ_2 (only) in (9·5) the following equations in the six unknowns a_2, a_4, b_3, b_4, c_2, c_3 were found:

Eqn. no.	Left: coefficients of						Right: absolute term
	a_2	a_4	b_3	b_4	c_2	c_3	
1	12	1	—	—	—	—	− 13·112496
2	0·816667	0·281315	0·287507	0·189756	—	—	− 0·403210
3	—	—	—	—	1·193685	1·304172	0·094626
4	1·897366	0·756233	1·306833	1·150028	2·185118	3·581055	− 0·122438
5	0·229605	0·069278	0·047439	0·025048	0·434724	0·356221	− 0·122152
6	0·130637	0·041871	0·032192	0·017835	0·668438	0·496634	− 0·044365

Approximations to F, G and H were found:

$$\left.\begin{aligned}
F_1(x) &\approx 0{\cdot}6781 - 1{\cdot}238888x^2 + 1{\cdot}754160x^4, \\
G_1(y) &\approx 0{\cdot}5870 - 0{\cdot}546478y - 0{\cdot}361808y^2 + 0{\cdot}171129y^3 + 0{\cdot}347163y^4, \\
H_1(z) &\approx 0{\cdot}078091z^{\frac{3}{2}} - 0{\cdot}017319z^2 + 0{\cdot}088408z^3.
\end{aligned}\right\} \qquad (9{\cdot}6)$$

These yielded estimates of the 4th and 6th moments as follows:

$$\mu_4' = 0{\cdot}419712, \quad \mu_6' = 0{\cdot}720776, \tag{9·7}$$

differing by $-0{\cdot}5\%$ and $-1{\cdot}7\%$ respectively from the correct values at (9·5). These deviations were not serious from the viewpoint of probability-point determination. Nevertheless, it seemed worth while to try to achieve a closer approximation. This was done by finding a 'corrector' $\phi(x)$ (not positive, like a frequency, for all values of x) with the following properties:

$$\left.\begin{array}{ll} \text{(i)} & \text{total 'frequency' zero,} \\ \text{(ii)} & \text{'2nd moment' zero,} \\ \text{(iii)} & \phi'(0) = 0, \\ \text{(iv)} & \phi(\gamma) = \phi'(\gamma) = 0, \\ \text{(v)} & \text{'4th moment'} = \mu_4 - \mu_4' = 0{\cdot}002163. \end{array}\right\} \tag{9·8}$$

Then $$\phi(x) = 0{\cdot}002404 - 0{\cdot}028853x^2 + 0{\cdot}045232x^3 - 0{\cdot}024236x^4 + 0{\cdot}004342x^5, \tag{9·9}$$

and the frequencies finally adopted are

$$\left.\begin{array}{ll} 0 - 1/\sqrt{12} \;\ldots & F(x) = F_1(x) + \phi(x), \\ 1/\sqrt{12} - 3/\sqrt{10} \;\ldots & G(y) = G_1(y) + \phi(x), \\ 3/\sqrt{10} - 5/\sqrt{6} \;\ldots & H(z) = H_1(z) + \phi(x), \end{array}\right\} \tag{9·10}$$

F_1, G_1 and H_1 being given by (9·6) and $x = t_7$. It is evident from the smallness of the coefficients of $\phi(x)$ in (9·9) that the correction effected by $\phi(x)$ is minute. From (9·10) the moment μ_6'' is 0·728972, so that the error is reduced to about one-third of what it was using F_1, G_1 and H_1.

10. Samples of 8

The links and link frequencies are as follows:

$$\left.\begin{array}{rcl} \text{Link} & & \text{Link frequency} \\ 0 & : & 0{\cdot}6927 = A, \\ \beta = 2/\sqrt{15} & : & 0{\cdot}4442 = B, \\ \gamma = 2/\sqrt{3} & : & 0{\cdot}1018 = C, \\ \delta = 6/\sqrt{7} & : & 0{\cdot}04019153(\delta - x)^2 = D(\delta - x)^2/2, \end{array}\right\} \tag{10·1}$$

where $x = t_8$.

Set

$$\left.\begin{array}{l} y = x - \beta, \\ z = \delta - x, \\ \kappa = \gamma - \beta = 0{\cdot}638303, \\ \lambda = \delta - \gamma = 1{\cdot}113086. \end{array}\right\} \tag{10·2}$$

The orders of contact (§3) entail the following forms for the three zones:

Zone

$$\left.\begin{array}{ll} 0 - 2/\sqrt{15}: & F(x) = A + a_2\dfrac{x^2}{2} + a_3\dfrac{x^3}{6} + a_4\dfrac{x^4}{24}, \\ 2/\sqrt{15} - 2/\sqrt{3}: & G(y) = B + \left(a_2\beta + a_3\dfrac{\beta^2}{2} + a_4\dfrac{\beta^3}{6}\right)y + \left(a_2 + a_3\beta + a_4\dfrac{\beta^2}{2}\right)\dfrac{y^2}{2} + b_3\dfrac{y^3}{6} + b_4\dfrac{y^4}{24}, \\ 2/\sqrt{3} - 6/\sqrt{7}: & H(z) = D\dfrac{z^2}{2} + c_3\dfrac{z^3}{6} + c_4\dfrac{z^4}{24}. \end{array}\right\} \tag{10·3}$$

Five of the seven equations required to determine the a, b and c will be found from the order of contact conditions, as follows:

(i) $B = A + a_2\frac{\beta^2}{2} + a_3\frac{\beta^3}{6} + a_4\frac{\beta^4}{24}$,

(ii) $C = B + \left(a_2\beta + a_3\frac{\beta^2}{2} + a_4\frac{\beta^3}{6}\right) + \left(a_2 + a_3\beta + a_4\frac{\beta^2}{2}\right)\frac{\kappa^2}{2} + b_3\frac{\kappa^3}{6} + b_4\frac{\kappa^4}{24}$,

(iii) $C = D\frac{\lambda^2}{2} + c_3\frac{\lambda^3}{6} + c_4\frac{\lambda^4}{24}$,

(iv) $a_2\beta + a_3\frac{\beta^2}{2} + a_4\frac{\beta^3}{6} + \left(a_2 + a_3\beta + a_4\frac{\beta^2}{2}\right)\kappa + b_3\frac{\kappa^2}{2} + b_4\frac{\kappa^3}{6} = -D\lambda - c_3\frac{\lambda^2}{2} - c_4\frac{\lambda^3}{6}$,

(v) $a_2 + a_3\beta + a_4\frac{\beta^2}{2} + b_3\kappa + b_4\frac{\kappa^2}{2} = D + c_3\lambda + c_4\frac{\lambda^2}{2}$.

The remaining two equations were found by equating the 0th and 2nd moments from the curves to the true values 1 and 4/11 respectively. The frequency functions found were as follows:

$$\left.\begin{aligned} F(x) &= 0\cdot6927 - 0\cdot320142x^2 - 2\cdot7751x^3 + 3\cdot08x^4, \\ G(y) &= 0\cdot4442 - 0\cdot854177y + 0\cdot308677y^2 + 0\cdot649680y^3 - 0\cdot553667y^4, \\ H(z) &= 0\cdot040192z^2 + 0\cdot027763z^3 + 0\cdot008933z^4, \end{aligned}\right\} \quad (10\cdot4)$$

where $x = t_8$.

For reasons which will be apparent in the next section, it was not deemed necessary to apply higher momental checks in this case.

Reference may here be made to yet another experiment, the negative result of which may have some interest. At the $n = 6$ stage the remarkable 'regularity' which the curve assumed, after its highly bizarre appearance at the stage before, suggested that orders of contact (except at the limit of range) might be ignored at a slightly later stage and a single curve fitted using the moments only.

Using μ_0, μ_2, μ_4 and μ_6, and the $D(\delta - z)^2/2$ (see (10·1)) for the forms at the limit of range with $F_1'(0) = 0$ the following frequency curve was found:

$$\begin{aligned} F_1(x) = 0\cdot040192(\delta-x)^2 + 0\cdot132866(\delta-x)^3 - 0\cdot293716(\delta-x)^4 + 0\cdot231146(\delta-x)^5 \\ - 0\cdot039279(\delta-x)^6 - 0\cdot005209(\delta-x)^7. \end{aligned} \quad (10\cdot5)$$

The correct values of the moments (to 6 places) were

$$\mu_0 = 1, \quad \mu_2 = 0\cdot363636, \quad \mu_4 = 0\cdot414644, \quad \mu_6 = 0\cdot763334, \quad \mu_8 = 1\cdot823617. \quad (10\cdot6)$$

The value μ_8' of the 8th moment computed from the curve was 1·993270, an error therefore of +9·3 %. The central ordinate $F_1(0) = 0\cdot9017$ as compared with the actual 0·6927, so that the curve $F_1(x)$ could not validly be used for further iteration, since the frequencies near the central frequency would be considerably in error. The probability (computed from (10·5)) for $F_1(x)$ beyond the 'true' 5 % probability point (computed from (10·3)) is 0·0456 which is quite accurate enough for practical purposes. This concordance, unexpected in view of the other facts mentioned, is due principally to the fact that $F_1(x)$ has the correct form at the limit of range. This experiment shows that, despite the regularity of the $\sqrt{b_1}$ distribution for $n = 8$, the problem of finding the nearly exact distribution cannot be treated in cavalier fashion.

11. PROBABILITY POINTS FOR FREQUENCIES FOR SAMPLES OF 8 OR MORE

By the Gram-Charlier theorem for symmetrical distributions under general conditions any frequency $f(w)$, where w has mean zero and variance unity, can be expanded in the form

$$f(w) = \exp\left\{\frac{\lambda_4}{4!\,\lambda_2^2}\left(\frac{d}{dw}\right)^4 + \frac{\lambda_6}{6!\,\lambda_2^3}\left(\frac{d}{dw}\right)^6 + \frac{\lambda_8}{8!\,\lambda_2^4}\left(\frac{d}{dw}\right)^8 + \dots\right\}\Theta(w), \tag{11·1}$$

where
$$\Theta(w) = \frac{1}{\sqrt{(2\pi)}}\exp - \tfrac{1}{2}w^2,$$

the λ being semi-invariants of the original variate. Let u be a normal variate with mean zero and variance unity. Using the method of E. A. Cornish & R. A. Fisher (1937) their expression for w in terms of u has been extended to the following effect:

$$w = u - \frac{\lambda_4}{24\lambda_2^2}x_3 - \frac{\lambda_4^2}{384\lambda_2^4}y_5 - \frac{\lambda_6}{720\lambda_2^3}x_5 + \frac{\lambda_4^3}{3072\lambda_2^6}y_7 - \frac{\lambda_4\lambda_6}{1152\lambda_2^5}z_7 - \frac{\lambda_8}{4032\lambda_2^4}x_7 + \dots, \tag{11·2}$$

where the x_k are Hermite polynomials in u of the degree indicated. The y_j and z_l terms in (11·2) are as follows:

$$\left.\begin{aligned} x_3 &= -u^3 + 3u, & y_7 &= 9u^7 - 131u^5 + 451u^3 - 321u,\\ y_5 &= 3u^5 - 24u^3 + 29u, & z_7 &= u^7 - 17u^5 + 69u^3 - 57u,\\ x_5 &= -u^5 + 10u^3 - 15u, & x_7 &= -u^7 + 21u^5 - 105u^3 + 105u. \end{aligned}\right\} \tag{11·3}$$

At (11·2) the expansion is taken to $O(n^{-3})$ because λ_{2k}/λ_2^k is $O(n^{-k+1})$ when the λ are semi-invariants of b_1 for samples of n.

The x_k, y_j and z_l functions at various probability levels are as follows:

Function	Probability points				
	0·10	0·05	0·025	0·01	0·001
u	1·281552	1·644854	1·959964	2·326348	3·090223
x_3	1·739867	0·484338	− 1·649229	− 5·610905	− 20·239354
y_5	− 2·97984	− 22·98240	−37·09056	− 30·28992	+ 226·9286
x_5	− 1·632248	7·789154	16·986942	22·868797	− 33·058481
y_7	136·1309	194·9563	−22·4505	−675·7597	−1286·263
z_7	19·09291	41·19959	27·20947	− 53·45639	− 261·9424
x_7	− 19·5234	− 74·2935	−88·4883	− 15·5752	362·6625

(11·4)

For $n = 8$ the semi-invariants, etc., required are

$$\left.\begin{aligned} \lambda_2 &= 0·363636, & \lambda_4/\lambda_2^2 &= 0·1357,\\ \lambda_4 &= 0·017950, & \lambda_6/\lambda_2^3 &= -1·1612,\\ \lambda_6 &= -0·055836, & \lambda_8/\lambda_2^4 &= 2·6577.\\ \lambda_8 &= 0·046470, \end{aligned}\right\} \tag{11·5}$$

If the formula at (11·2) were quite correct and then if we computed, at any probability level ϵ, the value of w, then set $x = \lambda_2^{\frac{1}{2}}w$ and from (10·4) computed the probability from end of range the result should be exactly ϵ, assuming, of course, that (10·4) gives the exact

frequency distribution. When this procedure is carried out at different pseudo-probability, i.e. the probability of x, levels indicated, the following results are found:

$$\left.\begin{array}{lccccc} \text{Pseudo-probability} & 0\cdot10 & 0\cdot05 & 0\cdot025 & 0\cdot01 & 0\cdot001 \\ (a)\ \text{True probability (to } x=\lambda_2^{\frac{1}{2}}w) & 0\cdot096855 & 0\cdot050459 & 0\cdot026825 & 0\cdot011504 & 0\cdot001155 \\ (b)\ \textit{Normal probability (to } x=\lambda_2^{\frac{1}{2}}u) & 0\cdot095564 & 0\cdot052376 & 0\cdot029502 & 0\cdot013419 & 0\cdot001090 \end{array}\right\} \quad (11\cdot6)$$

The correspondence at (a) is obviously satisfactory. At first sight it might appear that at 0·01 and 0·001 levels the divergence is (by the standards of this communication) rather marked. Actually this is not the case considering the fantastic difference in the algebraic form of the Gram-Charlier and the actual frequencies near the limit of range. The probabilities at (b) show that the normal curve gives quite a good representation. At $n=8$, however, the comparison flatters the normal curve since, as R. A. Fisher (1930) has shown, the ratio λ_4/λ_2^2 actually assumes its normal value of 3 at $n=7$ and reaches its greatest value at $n=22$.

We now propose to take a step which is discussed in some detail in the final section. We shall endow the right side of (11·2) with a remainder term which will make the probability of w formally the same as the pseudo-probability at (11·6). The following table shows the value of the variate $t_8\sigma^{-1}$ computed from (10·4) (where x represents t_8) at different true probability levels, together with the corresponding value of w computed from (11·2):

Probability	t_8/σ	w	R
0·10	1·253173	1·273231	−82·2
0·05	1·671682	1·666548	21·0
0·025	2·043181	2·008014	143·3
0·01	2·445125	2·389594	227·5
0·001	3·107977	3·079696	115·8

(11·7)

It has been seen that the difference between w as given by (11·2) and the true value is $O(n^{-4})$. Accordingly the values of R were found at the different probability levels by setting

$$w+\frac{R}{n^4}=\frac{t_8}{\sigma}, \tag{11·8}$$

with $n=8$. The estimates of the probability points P for values of $n \geqslant 8$ are accordingly

$$P=\lambda_2^{\frac{1}{2}}\left(A+B\frac{\lambda_4}{\lambda_2^2}+C\frac{\lambda_4^2}{\lambda_2^4}+D\frac{\lambda_6}{\lambda_2^3}+E\frac{\lambda_4^3}{\lambda_2^6}+F\frac{\lambda_4\lambda_6}{\lambda_2^5}+G\frac{\lambda_8}{\lambda_2^4}+\frac{R}{n^4}\right), \tag{11·9}$$

the values of $A, B, \ldots, G$ and R being given in the following table:

Probability	A	B	C	D	E	F	G	R
0·10	1·281552	−0·0724945	0·00776	0·00227	0·04431	−0·01657	0·000484	−82·2
0·05	1·644854	−0·0201808	0·05985	−0·01082	0·06346	−0·03576	0·001843	21·0
0·025	1·959964	0·0687179	0·09659	−0·02357	−0·00731	−0·02362	0·002195	143·3
0·01	2·326348	0·2337877	0·07888	−0·03176	−0·21997	0·04640	0·000386	227·5
0·001	3·090223	0·8433065	−0·59096	0·04591	−0·41871	0·22738	−0·008995	115·8

The terms in the first four columns agree with, or have been derived from Cornish & Fisher (1937). The λ's are semi-invariants derivable from the exact values of the moments given at (1·2).

As a test, the following is a comparison of the 0·05 and 0·01 probability points for $n = 25$ as derived by E. S. Pearson (1930) (using a Type VII curve) with the values from (11·9):

Probability level	Pearson	Geary (11·9)
0·05	0·711	0·707
0·01	1·061	1·062

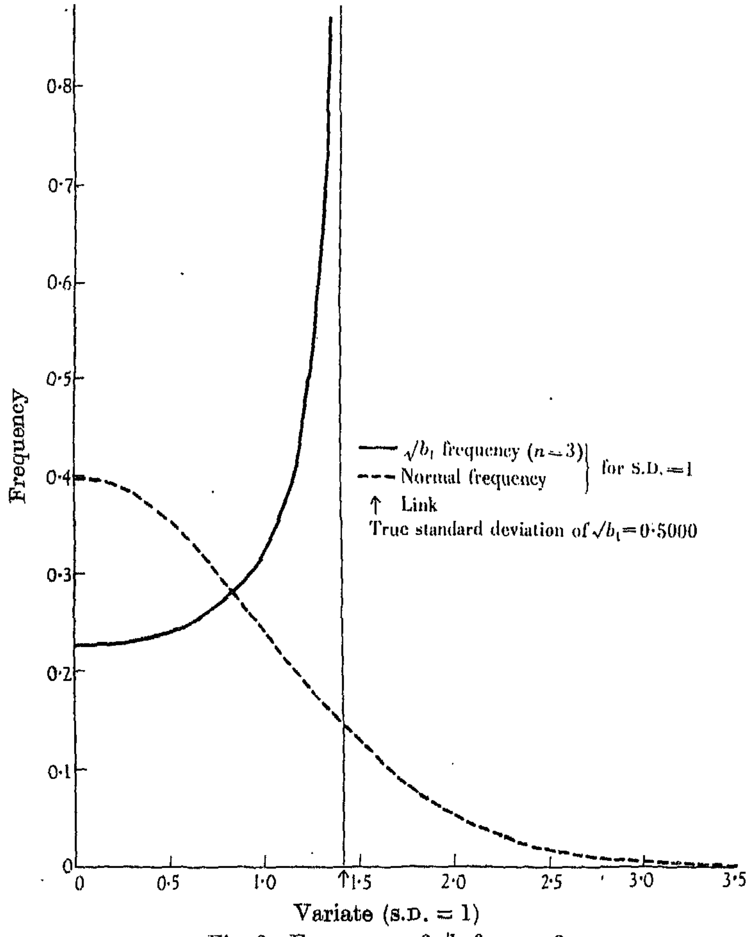


Fig. 3. Frequency of $\sqrt{b_1}$ for $n = 3$.

With standard deviation $\sigma = 0{\cdot}435$ it is obvious that the differences are not important. Sample number 25 is the lowest for which Pearson computed the probability points, and for two levels only. The formulae at (11·9) can probably be accepted with confidence.

12. Conclusion

From frequency formulae (5·2), (6·8), (7·25), (8·8), (9·10) (with (9·6) and (9·9)) and (10·4) the probability points for $\sqrt{b_1}$ for normal random samples of $n = 3, 4, 5, 6, 7$ and 8, respectively, can be determined without difficulty. The six frequency distributions are illustrated

in Figs. 3–8. On each of the $\sqrt{b_1}$ frequency curves there is superimposed the normal frequency with the same standard deviation, the intention being to enable a contrast to be made between the several $\sqrt{b_1}$ curves by reference each to the normal frequency, and to show the fairly rapid approach of the $\sqrt{b_1}$ frequency to normality with increasing n, even for small samples.*

In this research nothing was so remarkable as the transformation which the single step in the iteration, namely that from $n = 5$ to $n = 6$, effected in the shape of the frequency curve. From $n = 6$ on, the join at the links is effected so smoothly as to be almost imperceptible to the eye. The eye, however, flatters the actual approach to normality in the $\sqrt{b_1}$ frequency curves, as measured algebraically by the probability points.

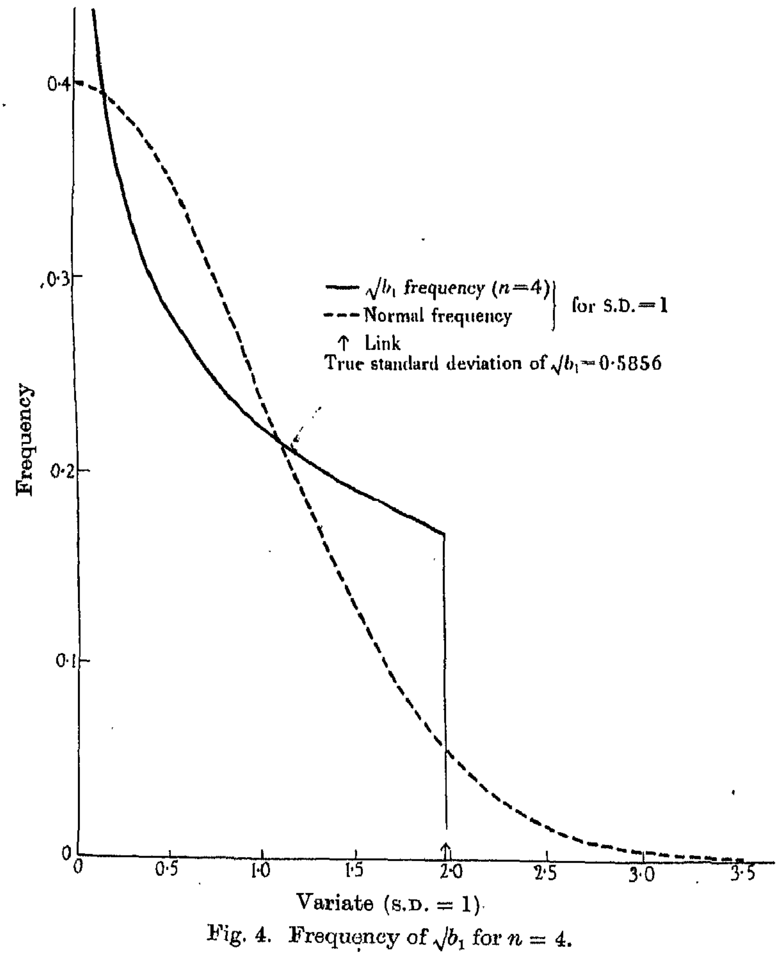


Fig. 4. Frequency of $\sqrt{b_1}$ for $n = 4$.

It may be well, at this stage, to recapitulate. Using integral iteration formula (2·13) (or (2·14)), frequency ordinates were computed at values of the variate termed the 'links' at which the frequency is shown to have functional discontinuities. Using the exact values of the moments (given at (1·2)), and taking into account the known order of contact (§ 3) of the different functions at the links and the known form assumed by the frequency at the

* As R. A. Fisher (1930) has shown, the approach to normality is not, however, uniform with increasing n, as indicated, say, by β_2. See p. 90 above.

known limit of range, inter-link frequencies were determined in polynomial form. Attention is directed to the use, at the $n = 4$ to 7 (inclusive) stages, of the higher moments for the purpose of checking the general reliability of the frequency curve (or rather series of curves joined at the links).

Of far greater practical importance, however, are the formulae (11·9) designed for the estimation of the 0·10, 0·05, 0·025, 0·01 and 0·001 probability points for normal random samples of $\sqrt{b_1}$ for $n \geqslant 8$. There will be little trouble about finding the corresponding formulae for other probability links. What degree of confidence can be reposed in these formulae? This raises in an acute form the vexed question (on which the protagonists of different schools were prone to get very vexed indeed a generation ago) of how best to use moments (or semi-invariants) for estimating frequency distributions. The general problem was constantly

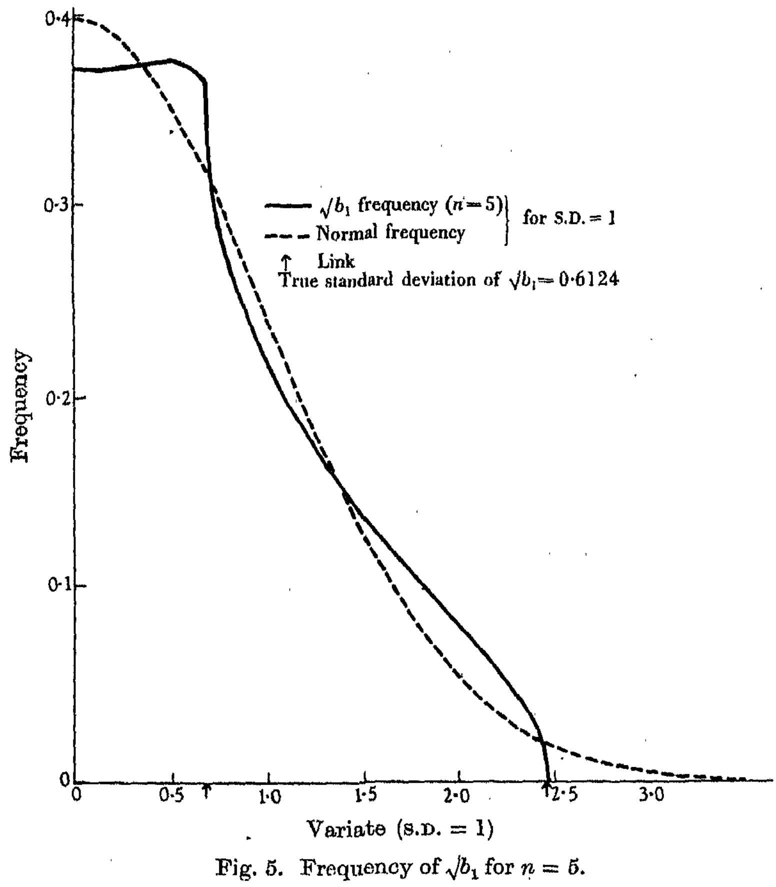


Fig. 5. Frequency of $\sqrt{b_1}$ for $n = 5$.

in the writer's mind during the present research and he would be glad if his colleagues could study the possibilities of the methods which culminated in formulae (11·9) for bridging the chasm which still divides the knowledge (sometimes exact) of the lower moments of statistics like $\sqrt{b_1}$ and b_2 and the formulae (however empirically established) for the frequency, in which a measure of confidence can be reposed. This fundamental problem was abandoned some years ago in a thoroughly unsatisfactory condition.

The Karl Pearson approach consists essentially in having regard to the 'shape' which experience has shown that frequency curves tend to assume and to use the first four moments

for the purpose of determining the constants of the curve. The disadvantage of the Pearson method is that of itself it gives no indication as to whether the resulting curve closely follows the actual frequency: it is necessary to have recourse to such devices as comparing the curve with a frequency distribution determined from hundreds of random sample computations of the statistic under examination. Apart from the tediousness of this method it is often indecisive in regard just to the parts of the frequency which are of most importance, namely the ends, because the small numbers which the check computation throws into these zones are usually subject to large (Poisson) errors.

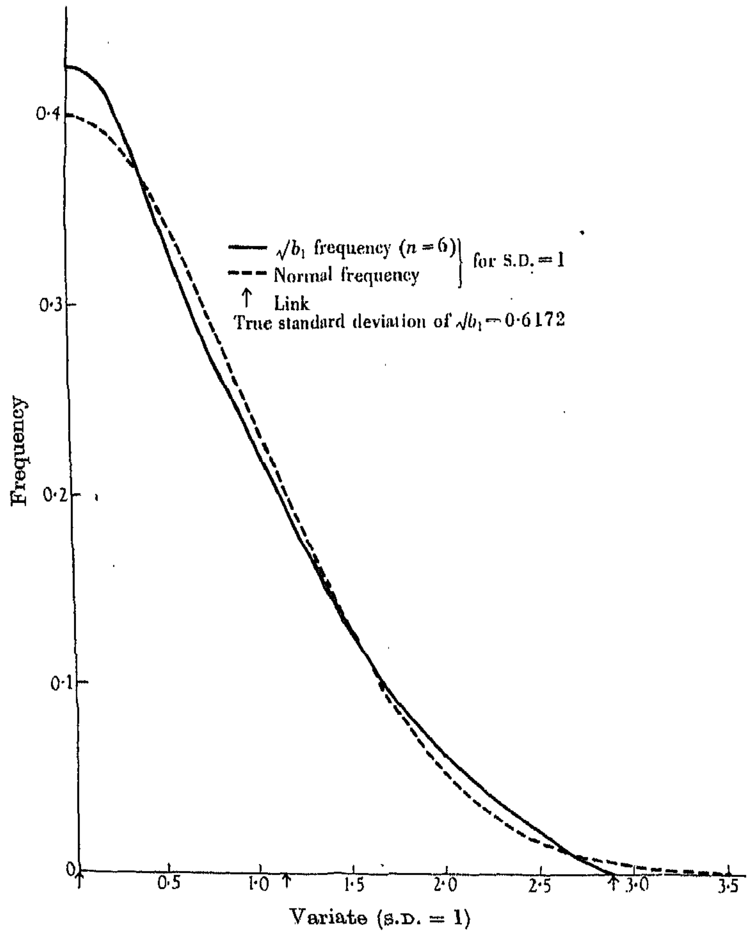


Fig. 6. Frequency of $\sqrt{b_1}$ for $n = 6$.

The Gram-Charlier system, on the other hand, can only be used with confidence when the frequency is fairly close to the normal. In practice the reliability is judged by the convergence of such terms as one can compute from the moments, i.e. if the successive terms show an 'unmistakable' tendency to diminish one feels confident in the computed frequency.

Obviously what both the Pearson, the Gram-Charlier and other frequency systems require is a Remainder Theorem. Since, however, an infinite number of moments are required to define a frequency distribution, with only a few moments known the most that can be expected is that upper (or lower) limits of the probability of the statistic can be established as

functions of the known moments. This is what Tchebychev's Theorem, and theorems of the type, do. Too much cannot be expected from the knowledge of a few moments: the approximations are almost invariably too rough for statistical use, when a high standard of efficiency is required; and M. Fréchet (1937) has shown that the Tchebychev type approximations are the best, given the assumptions, which can be made. For all their great *mathematical* importance (incidentally for their justification for the statistician of 'the faith that is in him'), it seems to the writer that research on these lines will not produce formulae which will be statistically utilizable in general conditions; but he may be quite wrong.

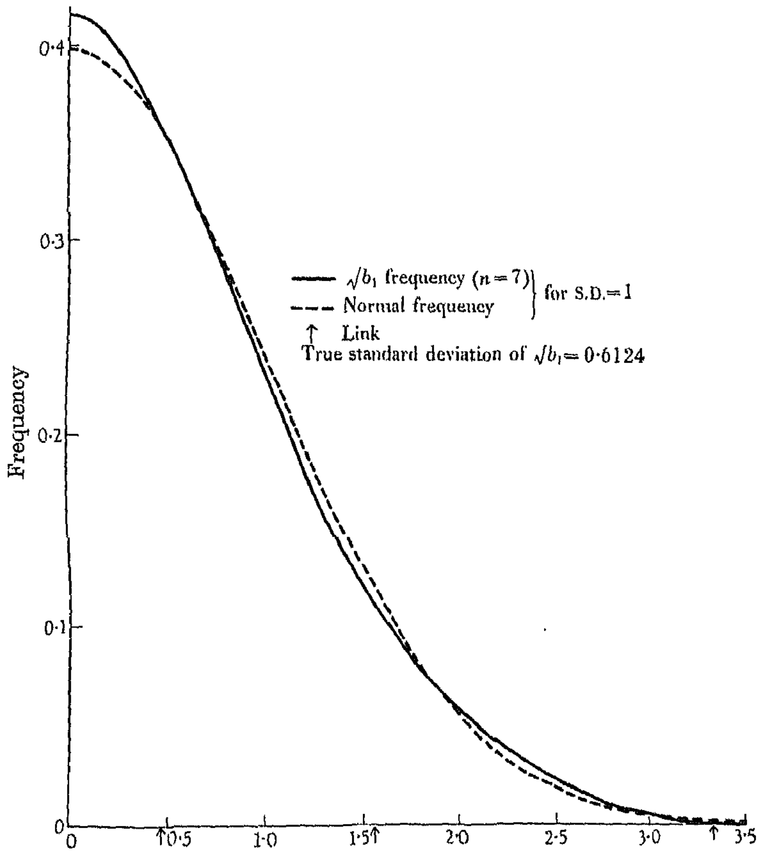


Fig. 7. Frequency of $\sqrt{b_1}$ for $n = 7$.

Knowing the earlier moments the Cornish-Fisher type expression (depending on the Gram-Charlier form of frequency) gives, at any probability level, an expansion for the variate to a defined order in the sample number. As might be surmised from the coefficients of the normal moments (e.g. (1·2) above), the coefficients in powers of $n^{-\frac{1}{2}}$ in the expansion of the variate usually tend to increase rapidly. In the present paper a remainder term of suitable order in n has been added to the known terms in the former expansion and its coefficient found by reference to the (assumed) exactly known expansion for $n = 8$. Clearly two more terms (in n^{-5} and n^{-6}) respectively could have been found had we iterated the frequency to $n = 9$ and $n = 10$, respectively, though this was not deemed necessary in the

ON THE COMPUTATION OF UNIVERSAL MOMENTS OF TESTS OF STATISTICAL NORMALITY DERIVED FROM SAMPLES DRAWN AT RANDOM FROM A NORMAL UNIVERSE. APPLICATION TO THE CALCULATION OF THE SEVENTH MOMENT OF b_2

By R. C. GEARY and J. P. G. WORLLEDGE

1. Introductory

The principal object of this communication is to develop a computational technique appropriate to the formula given by one of the authors (Geary, 1933). By way of illustration the formula is applied to the computation of the seventh moment of

$$b_2 = \frac{m_4}{m_2^2} = n \sum_{i=1}^{n} (x_i - \bar{x})^4 / \{\Sigma (x_i - \bar{x})^2\}^2, \tag{1·1}$$

where $x_1, x_2, \ldots x_n$ are the measures of the random sample of n and of which $\bar{x}$ is the arithmetic mean. Universal normality is assumed throughout.

A glance at formula (3·9) in which this paper culminates will indicate that the task of deriving higher normal moments of b_2 is not one to be undertaken in a frivolous spirit. The work finds its main justification in the conviction of the authors that accurate (if not exact) values of the probability points of b_2 can be found in terms of the moments of b_2 for all values of n using a method which has proved successful in the case of the analogous test of asymmetry, involving

$$\sqrt{b_1} = \frac{m_3}{m_2^{\frac{3}{2}}} = n^{\frac{1}{2}} \Sigma (x_i - \bar{x})^3 / \{\Sigma (x_i - \bar{x})^2\}^{\frac{3}{2}}. \tag{1·2}$$

In turn, the importance of the determination of accurate probabilities for $\sqrt{b_1}$ and b_2 for normal samples derives from the facts revealed by unpublished work by one of the authors. This shows (1) that probabilistic inferences drawn from the well-known significance tests based on the assumption of universal normality are apt to go astray when, in fact, the universe is not normal, and (2) that $\sqrt{b_1}$ and b_2 provide the most efficient tests of asymmetry and kurtosis, respectively, in indefinitely large samples, amongst wide fields of alternative tests and of alternative non-normal universes.

R. A. Fisher (1930) has given the exact values of the second, fourth and sixth moments of $\sqrt{b_1}$ and J. Pepper (1932) the eighth moment. In the former paper R. A. Fisher also gave the values of the second and third moments of b_2. The moment field was extended by J. Wishart (1930) and in a joint paper by R. A. Fisher & J. Wishart (1931). C. T. Hsu & D. N. Lawley (1940) gave the fifth and sixth moments of b_2. All these authors used the combinatorial method due to R. A. Fisher (1929). The present approach is entirely different.

2. The fundamental relation

To make the *exposé* complete it may be useful to reproduce the relevant part (which is quite brief) of the 1933 paper. The method used is due essentially to C. C. Craig (1928), applied to the normal case. Using, in the usual notation, a prefixed E to indicate 'expected' or, more

accurately, 'average value for all samples', we have for the characteristic function of the

$$z_i = x_i - \bar{x} \quad (i = 1, 2, \ldots, n), \tag{2·1}$$

the expression

$$E \exp \left\{ \sum_{i=1}^{n} t_i z_i \right\}, \tag{2·2}$$

where the t_i are n parameters, so that

$$E z_1^{a_1} z_2^{a_2} \ldots z_p^{a_p},$$

where the a_i are any p positive integers, is the coefficient of

$$t_1^{a_1} t_2^{a_2} \ldots t_p^{a_p}$$

in the expansion of

$$a_1! \, a_2! \ldots a_p! \, E \exp \{ t_1(x_1 - \bar{x}) + t_2(x_2 - \bar{x}) + \ldots + t_n(x_n - \bar{x}) \}. \tag{2·3}$$

The exponent can be written in the form

$$x_1(t_1 - \bar{t}) + x_2(t_2 - \bar{t}) + \ldots + x_n(t_n - \bar{t}),$$

where $\bar{t} = \sum_{i=1}^{n} t_i / n$. Since the x_i are independent,

$$E \exp \Sigma x_i(t_i - \bar{t}) = \prod_{i=1}^{n} E \exp x_i(t_i - \bar{t}). \tag{2·4}$$

Assuming, as we may without loss of generality, that the normal universe of the x_i has mean zero and unit standard deviation, we have

$$E \exp x_i(t_i - \bar{t}) = \exp \tfrac{1}{2}(t_i - \bar{t})^2.$$

Hence

$$a_1! \, a_2! \ldots a_p! \, E \exp \sum_{i=1}^{n} t_i(x_i - \bar{x}) = a_1! \, a_2! \ldots a_p! \exp \tfrac{1}{2} \Sigma (t_i - \bar{t})^2. \tag{2·5}$$

By definition, the *power* f of a term is given by $f = \sum_i a_i$ and the *dimension* by p. It is clear that the required universal mean value of

$$E(x_1 - \bar{x})^{a_1} \ldots (x_p - \bar{x})^{a_p}$$

will be found as the coefficient of $\Pi t_i^{a_i}$ in the expansion of

$$\frac{a_1! \, a_2! \ldots a_p!}{k! \, 2^k} \left\{ t_1^2 + t_2^2 + \ldots + t_p^2 - \frac{(t_1 + t_2 + \ldots + t_p)^2}{n} \right\}^k, \tag{2·6}$$

where $2k = f$.

3. THE COMPUTATIONAL SCHEME

The computational scheme, which is quite general, will most clearly be outlined by reference to the computation of the exact value of a specific moment (from origin) $\mu_7'(m_4)$, for the derivation of which it was primarily designed. Then

$$\mu_7'(m_4) = \frac{1}{n^7} E(z_1^4 + z_2^4 + \ldots + z_n^4)^7 = \frac{1}{n^7} \Big[nE(\cdot 28 \cdot) + n(n-1) \left\{ \frac{7!}{6!\,1!} E(\cdot 24 \cdot 4) \right.$$

$$\left. + \frac{7!}{5!\,2!} E(\cdot 20 \cdot 8) + \frac{7!}{4!\,3!} E(\cdot 16 \cdot 12 \cdot) \right\} + n(n-1)(n-2) \left\{ \frac{7!}{5!\,1!^2\,2!} E(\cdot 20 \cdot 4^2) \right.$$

$$\left. + \frac{7!}{4!\,2!\,1!} E(\cdot 16 \cdot 84) + \frac{7!}{3!^2\,2!\,1!} E(\cdot 12^2 \cdot 4) + \frac{7!}{3!\,2!^2\,2!} E(\cdot 12 \cdot 8^2) \right\}$$

$$+n(n-1)(n-2)(n-3)\left\{\frac{7!}{4!\,1!^3\,3!}E(\cdot 16\cdot 4^3)+\frac{7!}{3!\,2!\,1!^2\,2!}E(\cdot 12\cdot 84^2)+\frac{7!}{2!^3\,3!\,1!}E(8^3\,4)\right\}$$

$$+n(n-1)(n-2)(n-3)(n-4)\left\{\frac{7!}{3!\,1!^4\,4!}E(\cdot 12\cdot 4^4)+\frac{7!}{2!^2\,2!\,1!^3\,3!}E(8^2 4^3)\right\}$$

$$+n(n-1)(n-2)(n-3)(n-4)(n-5)\frac{7!}{2!\,1!^5\,5!}E(84^5)$$

$$+n(n-1)(n-2)(n-3)(n-4)(n-5)(n-6)\frac{7!}{1!^7\,7!}E(4^7)\Bigg], \tag{3·1}$$

where, for example,

$$E(\cdot 12\cdot 84^2) = Ez_1^{12}z_2^8z_3^4z_4^4 = E(x_1-\bar{x})^{12}(x_2-\bar{x})^8(x_3-\bar{x})^4(x_4-\bar{x})^4.$$

There are, accordingly, fifteen terms made up of one of dimension one, three of dimension two, four of dimension three, etc. The structure of the numerical coefficients will be noted: in particular that, when the power of a factorial appears in the denominator, its factorial also appears. Each of the fifteen E terms will be evaluated separately, grouped by dimensions and multiplied by the n-factors.

As already stated, the value of $E(a_1a_2a_3\ldots)$ will be found as the coefficient of

$$t_1^{a_1}t_2^{a_2}t_3^{a_3}\ldots$$

in the expansion of
$$\frac{a_1!\,a_2!\,a_3!\ldots}{14!\,2^{14}}\{\Sigma t_i^2+\nu(\Sigma t_i)^2\}^{14}, \tag{3·2}$$

with $\nu=-1/n$. In this case, of course, $f=\Sigma a_i=28$.

Expand (3·2) in powers of ν by the binomial theorem. Each of the ν power terms will, in general, make a numerical contribution to the value of $E(a_1a_2\ldots)$ which will, accordingly, be represented by a polynomial in ν of degree 14. The term in ν^s will be

$$\frac{a_1!\,a_2!\ldots}{14!\,2^{14}}\,\frac{14!\,\nu^s}{s!(14-s)!}(14-s)!\,(2s)!\sum_s\frac{1}{(a_1-2s_1)!\,s_1!(a_2-2s_2)!\,s_2!\ldots}. \tag{3·3}$$

In the Σ_s, summation extends to all non-negative integer series $s_1, s_2, \ldots$, so that $Ss_i=(14-s)$, s_1 being associated with a_1, s_2 with a_2, etc. The values which the s_i can assume are obviously restricted further by the condition that

$$a_i \geqslant 2s_i.$$

Let the series $(\Sigma_0, \Sigma_1, \Sigma_2, \ldots)$ be termed the *reciprocal factorial vector* (hereafter usually written 'r.f.v.') of $a_1, a_2, \ldots$, the terms of the vector being regarded as of the *order* indicated by the subscript. The vector will be indicated by clarendon type. From the computational point of view the following relation is fundamental:

$$\mathbf{A}\times\mathbf{B}=\mathbf{AB}, \tag{3·4}$$

where $\mathbf{A}=(a_1a_2\ldots)$ and $\mathbf{B}=(b_1b_2\ldots)$, and any other r.f.v. The multiplication sign at (3·4) is defined as follows: the terms of $\mathbf{A}$ are multiplied respectively by ν^0, ν^1, ν^2, etc., and added to give a scalar A; the terms of $\mathbf{B}$ *in the reverse order* are also multiplied respectively by $\nu^0, \nu^1, \nu^2, \ldots$ and summed to give B. The coefficients of $\nu^0, \nu^1, \nu^2, \ldots$ in the product (in the ordinary sense) AB give the vector $\mathbf{AB}$. Relation (3·4) is immediately evident from the form of Σ_s in (3·3). From this relation it is quite easy to build up r.f.v.'s from those of lower order **44** from **4**, **84** from **8** and **4**, **88444** from **8** and **8444**, or **88** and **444**, etc.

Having found all fifteen r.f.v.'s the second step in the computational process is to form the scalar product of each r.f.v. and $(2s)!\,\nu^s/s!$—the latter will be termed the *ν-multipliers* which, it is important to note, are the same for all the terms in (3·1)—which, from (3·3), gives $E(a_1 a_2 \ldots)$ divided by

$$a_1!\,a_2!\ldots/2^{14}.$$

The latter are multiplied by the numerical factors in (3·1) to give what are termed the *constant multipliers*, 'constant' in the sense that they are the same for all the ν power terms in each of the E's in (3·1), but these constant terms are different for the different E's. For example, the constant multiplier for the term $Ez_1^{12} z_2^8 z_3^4 z_4^4$ is

$$\frac{12!\,8!\,4!^2\,7!}{3!\,2!\,1!^2\,2!\,2^{14}}. \tag{3·5}$$

Note the 'absolute constants' $7!$ and 2^{14}, and that the powers of the term appear as factorials in the numerator and factorials one-fourth of these powers in the denominator. In the denominator is also a $2!$ which is the factorial of the factorial power.

The third step in the computation is to sum the terms of the same dimensions. The final step consists in the multiplication of the terms of the different dimensions by the *ν-factors* as follows:

Table 1. *ν-factors*

Dimension	ν-factors
1	ν^6
2	$-(\nu^6+\nu^5)$
3	$2\nu^6+3\nu^5+\nu^4$
4	$-(6\nu^6+11\nu^5+6\nu^4+\nu^3)$
5	$24\nu^6+50\nu^5+35\nu^4+10\nu^3+\nu^2$
6	$-(120\nu^6+274\nu^5+225\nu^4+85\nu^3+15\nu^2+\nu)$
7	$720\nu^6+1764\nu^5+1624\nu^4+735\nu^3+175\nu^2+21\nu+1$

(3·6)

The ν-factors at (3·6) are, of course, the n-factors in (3·1) with $\nu = -1/n$.

To deal with the very large whole numbers and their reciprocals which arise in factorial computation we had recourse to a *prime number index* notation. For this purpose the number is factorized into powers of the lower primes—we have used the notation for primes not exceeding 31. Thus

$$746{,}137{,}199{,}808{,}000 = 6847\cdot 13^1\cdot 11^1\cdot 7^2\cdot 5^3\cdot 3^5\cdot 2^9$$

is written in this notation $\qquad$ 6847[112359],

the digits in the square brackets [ ] being the powers of the lowest primes arranged in ascending order from the right. The ordinary number 6847 will be known as the *coefficient* and the symbolical number in square brackets as the *primal* of the original number. Note that in this example the notation affects an economy from 15 to 10 in the number of digits required to describe the number. Should the original number not be factorizable by a particular small prime a 0 will be inserted in the proper place, e.g. [10358] means that 7 is not a factor of the number represented. If, as often happens with the first two primes, the indices exceed 9, decimal points are used, e.g. [124·11·17] means that the original number has 2^{17} and 3^{11} as factors. The primal notation can be used when the indices are all positive

or all negative: occasionally, however, + and − signs have to be mixed in the primal (see Table 6).

With little practice great facility is acquired in applying the ordinary rules to numbers in primal notation. For multiplication or division corresponding digits in the primals are added or subtracted, the coefficients being dealt with in the ordinary way. In addition or subtraction common factors in the primals are immediately evident and the coefficient of the sum (or difference) is derived usually by a single product-sum (or product-difference) operation on a multiplying machine. It may be observed that all the work for this paper was executed without inconvenience on small hand multiplying machines with capacity $9 \times 8 \times 13$.

In the following tables the first thirty-two factorials, the ν-multipliers and the constant multipliers required for the computation of $\mu_7'(m_4)$ are expressed in primal notation.

Table 2. *Factorials in primal notation*

0! = 1! =	[0]	17! =	[111236·15]
2! =	[1]	18! =	[111238·16]
3! =	[11]	19! =	[1111238·16]
4! =	[13]	20! =	[1111248·18]
5! =	[113]	21! =	[1111349·18]
6! =	[124]	22! =	[1112349·19]
7! =	[1124]	23! =	[11112349·19]
8! =	[1127]	24! =	[1111234·10·22]
9! =	[1147]	25! =	[1111236·10·22]
10! =	[1248]	26! =	[1112236·10·23]
11! =	[11248]	27! =	[1112236·13·23]
12! =	[1125·10]	28! =	[1112246·13·25]
13! =	[11125·10]	29! =	[11112246·13·25]
14! =	[11225·11]	30! =	[11112247·14·26]
15! =	[11236·11]	31! =	[111112247·14·26]
16! =	[11236·15]	32! =	[111112247·14·31]

Table 3. *ν-Multipliers in factorial and primal notation*

Term in			Coefficient
ν^0	:	$0!\ 0!^{-1} =$	[0]
ν^1	:	$2!\ 1!^{-1} =$	[1]
ν^2	:	$4!\ 2!^{-1} =$	[12]
ν^3	:	$6!\ 3!^{-1} =$	[113]
ν^4	:	$8!\ 4!^{-1} =$	[1114]
ν^5	:	$10!\ 5!^{-1} =$	[1135]
ν^6	:	$12!\ 6!^{-1} =$	[11136]
ν^7	:	$14!\ 7!^{-1} =$	[111137]
ν^8	:	$16!\ 8!^{-1} =$	[111248]
ν^9	:	$18!\ 9!^{-1} =$	[1111249]
ν^{10}	:	$20!\,10!^{-1} =$	[1111124·10]
ν^{11}	:	$22!\,11!^{-1} =$	[1111225·11]
ν^{12}	:	$24!\,12!^{-1} =$	[11111225·12]
ν^{13}	:	$26!\,13!^{-1} =$	[11111245·13]
ν^{14}	:	$28!\,14!^{-1} =$	[11111248·14]

Table 4. *Constant multipliers in factorial and primal notation*

Required for computation of the undermentioned term in (3·1)				
$E(28)$	:	$28!\,7!/7!\,2^{14}$	=	[1112246·13·11]
$E(24\cdot 4)$	:	$24!\,4!\,7!/6!\,1!\,2^{14}$	=	[1111244·11·11]
$E(20\cdot 8)$	:	$20!\,8!\,7!/5!\,2!\,2^{14}$	=	[111145·11·11]
$E(16\cdot 12)$	:	$16!\,12!\,7!/4!\,3!\,2^{14}$	=	[1246·11·11]
$E(20\cdot 4^2)$	:	$20!\,4!^2\,7!/5!\,1!^2\,2!\,2^{14}$	=	[111134·11·10]
$E(16\cdot 84)$	:	$16!\,8!\,4!\,7!/4!\,2!\,1!\,2^{14}$	=	[1145·10·11]
$E(12^2\,4)$	:	$12!^2\,4!\,7!/3!^2\,2!\,1!\,2^{14}$	=	[235·11·10]
$E(12\cdot 8^2)$	:	$12!\,8!^2\,7!/3!\,2!^2\,2!\,2^{14}$	=	[145·10·10]
$E(16\cdot 4^3)$	:	$16!\,4!^3\,7!/4!\,1!^3\,3!\,2^{14}$	=	[1134·9·10]
$E(12\cdot 84^2)$	:	$12!\,8!\,4!^2\,7!/3!\,2!\,1!^2\,2!\,2^{14}$	=	[134·10·10]
$E(8^3\,4)$	:	$8!^3\,4!\,7!/2!^3\,3!\,1!\,2^{14}$	=	[44·8·10]
$E(12\cdot 4^4)$	:	$12!\,4!^4\,7!/3!\,1!^4\,4!\,2^{14}$	=	[12398]
$E(8^2\,4^3)$	:	$8!^2\,4!^3\,7!/2!^2\,2!\,1!^3\,3!\,2^{14}$	=	[3389]
$E(84^5)$	:	$8!\,4!^5\,7!/2!\,1!^5\,5!\,2^{14}$	=	[2188]
$E(4^7)$	:	$4!^7\,7!/1!^7\,7!\,2^{14}$	=	[77]

The theory will be illustrated by reference to the computation of $Ez_1^8 z_2^8 z_3^4 z_4^4 z_5^4 = E(8^2\,4^3)$. First the r.f.v. **88444** is found as the product **884** × **44** by setting down in equal spaces the terms of **884** and on a movable slip spaced to the former the terms of **44** in reverse:

All primals are negative

884	[27]	5[27]	109[39]	111[137]	803[112·10]	1493[114·10]	389[122·12]	119[124·11]	1543[225·15]	31[225·15]	[225·17]

MOVABLE SLIP →						
	[26]	[13]	7[13]	[1]	[2]	44

The term in **88444** from the position illustrated is that of the 5th order, namely,

$$5[\underline{4}\cdot\underline{13}]+109[\underline{4}\cdot 12]+111\cdot 7[\underline{14}\cdot 10]+803[\underline{112}\cdot 11]+1493[\underline{114}\cdot 12]$$

$$=[114\cdot 13]\,(5\cdot 7\cdot 5+109\cdot 7\cdot 5\cdot 2+777\cdot 7\cdot 8+803\cdot 9\cdot 4+1493\cdot 2)$$

$$=[114\cdot 13]\,(3\cdot 27737)=27737[113\cdot 13].$$

The manner of computation is indicated: first the largest (negative) digits in each of the four positions of the primals are underlined and the underlined set is regarded as the common factor. Note how, at the final stage, the factor 3 of the coefficient reduces the primal digit from 4 to 3. From the entries in the round brackets ( ) it will be clear that, as stated above, the procedure is well adapted to the multiplying machine. The full calculation of **88444** is shown in Table 5.

The identity of the r.f.v.'s from the two factorizations of **88444** constitutes an absolute check of the work. The calculation of $E(8^2\,4^3)$ required for (3·1) is completed in Table 6. In practice the figures in columns (4) and (5) of this table were derived from those in column (3), and in Tables 4 and 5 by entering the latter on two movable slips and folding opposite each entry, as required. This stage of the work was rapidly executed. The sum-product of columns (1), (2) and (5) give the value of $E(8^2\,4^3)$. All the r.f.v.'s required for the calculation of the E's for (3·1) are given in the appendix.

Table 5. *Calculation of reciprocal factorial vector* **88444**

All primals are negative

(i) By 884×44

Order			r.f.v. of 88444
0 :	[29]	=	[29]
1 :	[2·8]+5[29]	=	7[29]
2 :	7[3·10]+5[28]+109[3·11]	=	9[·11]
3 :	[3·10]+35[3·10]+109[3·10]+111[139]	=	947[13·10]
4 :	[4·13]+5[3·10]+763[4·12]+111[138]+803[112·12]	=	1811[110·13]
5 :	5[4·13]+109[4·12]+777[14·10]+803[112·11]+1493[114·12]	=	27737[113·13]
6 :	109[5·15]+111[14·10]+5621[113·13]+1493[114·11]+389[122·14]	=	1783141[125·15]
7 :	111[15·13]+803[113·13]+1493[15·13]+389[122·13]+119[124·13]	=	20627[115·13]
8 :	803[114·16]+1493[115·13]+389[23·15]+119[124·12]+1543[225·17]	=	1772417[225·17]
9 :	1493[116·16]+389[123·15]+119[25·14]+1543[225·16]+31[225·17]	=	547889[226·17]
10 :	389[124·18]+119[125·14]+1543[126·18]+31[225·16]+[225·19]	=	151331[226·19]
11 :	119[126·17]+1543[226·18]+217[226·18]+[225·18]	=	127[223·18]
12 :	1543[227·21]+31[226·18]+[126·20]	=	2329[227·21]
13 :	31[227·21]+[226·20]	=	37[227·21]
14 :	[227·23]	=	[227·23]

(ii) By 8844×4

Order			r.f.v. of 88444
0 :	[29]	=	[29]
1 :	[29]+[18]	=	7[29]
2 :	[3·11]+[18]+85[3·10]	=	9[·11]
3 :	[2·10]+85[3·10]+169[12·10]	=	947[13·10]
4 :	85[4·12]+169[12·10]+11113[104·13]	=	1811[110·13]
5 :	169[13·12]+11113[104·13]+5137[114·11]	=	27737[113·13]
6 :	11113[105·15]+5137[114·11]+22703[124·13]	=	1783141[125·15]
7 :	5137[115·13]+22703[124·13]+9341[125·13]	=	20627[115·13]
8 :	22703[125·15]+9341[125·13]+90541[225·17]	=	1772417[225·17]
9 :	9341[126·15]+90541[225·17]+2453[225·16]	=	547889[226·17]
10 :	90541[226·19]+2453[225·16]+137[126·18]	=	151331[226·19]
11 :	2453[226·18]+137[126·18]+17[226·18]	=	127[223·18]
12 :	137[127·20]+17[226·18]+[226·21]	=	2329[227·21]
13 :	17[227·20]+[226·21]	=	37[227·21]
14 :	[227·23]	=	[227·23]

Finally, the E's are multiplied by the appropriate ν-factors given in Table 1, to give the value of $E(m_4^7)$. Now R. A. Fisher (1930) (see also Geary, 1933) has shown that

$$\mu_7'(b_2) = E(b_2^7) = E(m_4^7)/E(m_2^{14}) \tag{3·7}$$

and
$$E(m_2^{14}) = (n-1)(n+1)(n+3)\dots(n+23)(n+25)/n^{14}. \tag{3·8}$$

Finally,
$$\begin{aligned}\mu_7'(b_2) = (&3^7n^{13} + 211\cdot 3^7n^{12} + 64{,}802\cdot 3^6n^{12} + 13{,}154{,}290\cdot 3^5n^{10} \\ &+ 668{,}584{,}331\cdot 3^6n^9 + 25{,}489{,}306{,}481\cdot 3^5n^8 + 74{,}020{,}784{,}452\cdot 7\cdot 3^5n^7 \\ &- 72{,}634{,}851{,}124\cdot 7\cdot 5\cdot 3^6n^6 + 407{,}081{,}273{,}655\cdot 7\cdot 5\cdot 3^6n^5 \\ &- 1{,}287{,}510{,}783{,}723\cdot 7\cdot 5\cdot 3^6n^4 + 2{,}526{,}463{,}322{,}982\cdot 7\cdot 5\cdot 3^6n^3 \\ &- 280{,}521{,}238{,}122\cdot 11\cdot 7\cdot 5\cdot 3^6n^2 + 3{,}036{,}544{,}767\cdot 13\cdot 11\cdot 7\cdot 5^2\cdot 3^6n \\ &- 135{,}393{,}525\cdot 13\cdot 11\cdot 7^2 5^2 3^6)/(n+1)(n+3)\dots(n+23)(n+25).\end{aligned} \tag{3·9}$$

4. CORROBORATION OF FORMULAE

An integral part of the present work is the technique of check. To be of value the formulae at (3·9) and (3·10) must be absolutely correct because (1) any errors made in factorial work are fairly certain to be large and (2) the formulae are designed for use when n is small, when relatively small errors in the numerical coefficients may materially affect the results. Furthermore, it is almost impossible to avoid error (even in a joint work like the present) with so many individual calculations involving numbers astronomically large. As will appear, there is a satisfactory, though not absolute, check at the final stage; but if it reveals error it does not show where the error occurred, so that, if this were the sole check, there would be no

Table 6. *Calculation of* $E(8^2 4^3)$ *from* **88444**

Term (1)	r.f.v. 88444		(3) × ν-multiplier (Table 3) (4)	(4) × constant multiplier [3389] (5)
	Coefficient (2)	Primal (neg.) (3)		
ν^0	1	[29]	[−2−9]	[3360]
ν^1	7	[29]	[−2−8]	[3361]
ν^2	9	[·11]	[1−9]	[3390]
ν^3	947	[13·10]	[−2−7]	[3362]
ν^4	1811	[110·13]	[1−9]	[3390]
ν^5	27737	[113·13]	[−8]	[3381]
ν^6	1783141	[125·15]	[10−1−2−9]	[13260]
ν^7	20627	[115·13]	[1100−2−6]	[113363]
ν^8	1772417	[225·17]	[11−10−1−9]	[112370]
ν^9	547889	[226·17]	[111−10−2−8]	[1112361]
ν^{10}	151331	[226·19]	[1111−10−2−9]	[11112360]
ν^{11}	127	[223·18]	[1111002−7]	[111133·10·2]
ν^{12}	2329	[227·21]	[1111100−2−9]	[111113360]
ν^{13}	37	[227·21]	[1111102−2−8]	[111113561]
ν^{14}	1	[227·23]	[11111021−9]	[111113590]

alternative but to face the tedium of complete recalculation. It is essential to devise an absolute check *at each stage*. This has been done for the present technique.

The first check is the $n = 1$ (or $\nu = -1$) check. This is applicable to the E's (see (3·1)) of dimension one, two and three. It derives from the fact that

$$\left\{\Sigma t_i^2 - \frac{1}{n}(\Sigma t_i)^2\right\}^{14} \equiv \left\{(1+\nu)\,\Sigma t_i^2 + 2\nu \sum_{i>j}\sum t_i t_j\right\}^{14}. \tag{4·1}$$

It will be immediately evident from the latter that when $\nu = -1$ the following terms vanish identically:

(i) all terms of one dimension;

(ii) all terms of two dimensions except those of the type $t_1^{14} t_2^{14}$ with which we are not concerned;

(iii) all terms of three or four dimensions in which the highest power exceeds 14: the latter being the highest power which, say, t_1 can assume in the expansion of

$$2^{14}\Big(\sum_{i>j}\sum t_i t_j\Big)^{14}\nu^{14}.$$

Even in terms of three dimensions in which the highest power is less than 14, e.g. in $Ez_1^{12}z_2^{12}z_3^4$, the $\nu = -1$ test can be exploited. In fact, from (2·5) and (4·1) the required terms for $\nu = -1$ are

$$E(12^2\cdot 4) = \frac{14!}{10!\,2!^2}\frac{12!^2\,4!}{3!^2\,2!}\frac{7!}{14!\,2^{14}}2^{14} = 12![11124],$$

$$E(12\cdot 8^2) = \frac{14!}{6!^2\,2!}\frac{12!\,8!^2}{3!\,2!^2\,2!}\frac{7!}{14!\,2^{14}}2^{14} = 12![3115].$$

The sum of these two terms is 131[1236·14] which should be the sum of the four E terms of dimension three in (3·1) since $E(20\cdot 4^2)$ and $E(16\cdot 84)$ are zero (for $\nu = -1$). The checks specified in this paragraph were fully applied to the terms of one, two and three dimensions in (3·1) before multiplication by the ν-factors (Table 1).

Reciprocal factorial vectors for dimensions exceeding two were checked fully by the 'double' factorization technique exemplified in Table 5. In view of the simplicity of the two subsequent processes, namely those of the ν- and constant multipliers, this check may be taken as establishing the accuracy of the E's of dimension three or more. Reference may nevertheless be made to a check at this stage, namely that the ratios of consecutive coefficients in each E exhibit a marked regularity, if correct. Any irregularity (which in the nature of the work will usually be large) must be suspect.

Assuming the accuracy of the E's in (3·1) the final stage was checked by multiplying by the ν-factors (3·6) in two ways:

(i) by straight multiplication using the primal notation;

(ii) by taking (in (3·1))

$$= (1+\nu)(1+2\nu)\ldots(1+6\nu)A_1 + \nu(1+\nu)\ldots(1+5\nu)A_2 + \ldots + \nu^6 A_7$$

$$= (1+\nu)\ldots(1+5\nu)\{(1+6\nu)A_1 + \nu A_2\} + \ldots,$$

and computing in successive stages

$$B_2 = (1+6\nu)A_1 + \nu A_2, \quad B_3 = (1+5\nu)B_2 + \nu^2 A_3, \text{ etc.}$$

The results were the same.

A satisfactory check for the final stage is that of $n = 4$. A. T. McKay (1933) has, in fact, given a formula for this value of n from which the seventh moment from zero of b_2 is found to be

$$82{,}220{,}810{,}251/5^2 11\cdot 13\cdot 17\cdot 23\cdot 29,$$

which value also transpired on substituting 4 for n in (3·9). This establishes the accuracy of all the formula except possibly the part accruing from the terms in (3·1) in

$$n(n-1)\ldots(n-4), \quad n(n-1)\ldots(n-5) \quad \text{and} \quad n(n-1)\ldots(n-6)$$

which vanish when $n = 4$.

If it adds nothing to the check in the previous paragraph it is nevertheless of interest to observe that, for $n = 3$, the value of the seventh moment of b_2 is found to be $(\frac{3}{2})^7$ which is as

it should be since, in this case, each b_2 assumes the constant value $\frac{3}{2}$, whether the samples are normal or not.

A partial check is also afforded at the final stage by the vanishing of all coefficients of powers of ν from ν^{15} to ν^{20} inclusive.

5. CONCLUSION

Previous investigators in this field have all used the combinatorial technique, invented by R. A. Fisher (1929) and applied in the first instance to the cumulants, which are linear functions of the sample moments. The present writers have not had sufficient experience in working the Fisher technique to decide which method is easier to apply. It is quite likely that the Fisher method is shorter. A strong point of the present computational scheme is that it lends itself to check at every stage; and the method may appeal to students who prefer the algebraical or arithmetical to the geometrical approach. For their benefit, and also in case it may later be found necessary (in connexion with the accurate determination of the probability points of b_2 for samples of all sizes) to compute higher moments than the seventh—it is almost certain the seventh will be required—we give as an appendix an extended series of reciprocal factorial vectors. From these can be derived without difficulty (i) corresponding E's, e.g. $E(x_1-\bar{x})^8(x_2-\bar{x})^4(x_3-\bar{x})^4$, on multiplication by appropriate ν- and constant multipliers and (ii) r.f.v.'s of higher powers.

APPENDIX

A selection of reciprocal factorial vectors required for the calculation of moments of b_2 for normal samples, including all used for the calculation of the seventh moment, in primal notation

All primals are negative

Order	4	8	4^2	12	84
0	[1]	[13]	[2]	[124]	[14]
1	[1]	[12]	[1]	[114]	[4]
2	[13]	[14]	7[13]	[26]	31[26]
3		[124]	[13]	[135]	7[115]
4		[1127]	[26]	[1128]	127[1128]
5				[1148]	17[1138]
6				[1125·10]	[113·10]

	4^3	16	12·4	8^2	84^2
0	[3]	[1127]	[125]	[26]	[15]
1	3[3]	[1125]	[123]	[24]	[13]
2	13[5]	[137]	13[135]	5[26]	17[25]
3	3[4]	[237]	[35]	31[136]	53[125]
4	13[17]	[113·10]	47[1138]	323[1139]	2497[1138]
5	[17]	[1259]	29[1247]	[1028]	173[1137]
6	[39]	[1125·11]	157[11259]	43[124·10]	7[139]
7		[11225·11]	[11249]	[124·10]	[1049]
8		[11236·15]	[1126·13]	[224·14]	[114·13]

Order	4^4	20	$16\cdot 4$	$12\cdot 8$	$12\cdot 4^2$
0	[4]	[1248]	[1128]	[137]	[126]
1	[2]	[1148]	[1028]	[37]	[26]
2	19[14]	[113·10]	11[13·10]	31[139]	101[138]
3	5[4]	[124·8]	47[1238]	7[227]	19[136]
4	49[17]	[124·11]	253[124·11]	299[123·10]	1163[1149]
5	5[16]	[135·11]	89[125·11]	73[115·10]	53[239]
6	19[38]	[1126·13]	51[1115·13]	3713[1126·12]	629[1105·11]
7	[38]	[11226·12]	1277[11226·12]	83[1126·11]	479[1125·10]
8	[4·12]	[11236·16]	229[11234·16]	1181[1236·15]	773[1126·14]
9		[111238·16]	[11135·16]	47[1237·15]	13[1126·14]
10		[1111248·18]	[11237·18]	[1237·17]	[1127·16]

	$8^2 4$	84^3	4^5	24	$20\cdot 4$
0	[27]	[16]	[5]	[1125·10]	[1249]
1	5[27]	5[16]	5[5]	[11249]	[1238]
2	109[39]	13[8]	125[17]	[125·11]	53[125·10]
3	111[137]	143[126]	35[15]	[126·11]	31[125·10]
4	803[112·10]	1399[1119]	545[28]	[224·14]	23[124·13]
5	1493[114·10]	239[1029]	23[8]	[236·12]	13[135·11]
6	389[122·12]	6943[114·11]	545[3·10]	[1137·14]	151[1136·13]
7	119[124·11]	65[104·10]	35[39]	[11236·14]	733[11236·13]
8	1543[225·15]	277[114·14]	125[4·13]	[11237·18]	1153[11237·17]
9	31[225·15]	23[115·14]	5[4·13]	[111239·17]	587[111238·16]
10	[225·17]	[115·16]	[5·15]	[1111248·19]	67[1101247·18]
11				[1112349·19]	[1111049·18]
12				[1111234·10·22]	[1111249·21]

	$16\cdot 8$	28	12^2	8^3	$8^2 4^2$
0	[113·10]	[11225·11]	[248]	[39]	[28]
1	[1129]	[11125·11]	[237]	[28]	[17]
2	13[104·11]	[1126·13]	23[249]	[·10·]	85[39]
3	43[114·11]	[1136·12]	47[259]	47[12·10]	169[129]
4	3823[224·14]	[236·15]	289[124·12]	89[101·13]	11113[104·12]
5	1507[226·12]	[238·15]	593[136·10]	281[113·11]	5137[114·10]
6	4933[1136·14]	[1237·17]	8531[1137·12]	2833[124·13]	22703[124·12]
7	28943[11236·14]	[11337·15]	193[1136·12]	11[121·13]	9341[125·12]
8	4331[1237·18]	[11248·19]	929[1236·16]	2213[224·17]	90541[225·16]
9	79[11235·17]	[111249·19]	113[1238·15]	2089[236·16]	2453[225·15]
10	43[11237·19]	[1111249·21]	37[1248·17]	71[235·18]	137[126·17]
11	37[11348·19]	[111234·10·20]	[1249·17]	[235·18]	17[226·17]
12	[11348·22]	[1111234·10·23]	[224·10·20]	[336·21]	[226·20]
13		[1112236·10·23]			
14		[1112246·13·25]			

Order	24·4	20·8	$20\cdot4^2$	16·12
0	[1125·11]	[125·11]	[124·10]	[124·11]
1	[1025·11]	[25·11]	[24·10]	[24·11]
2	139[1126·13]	19[124·13]	179[125·12]	187[125·13]
3	47[1126·12]	331[136·12]	87[125·11]	379[135·12]
4	79[226·15]	7001[236·15]	151[26·14]	2819[234·15]
5	31[137·15]	1489[236·15]	1501[136·14]	17161[237·15]
6	1051[1237·17]	25513[1237·17]	1609[1036·16]	104507[1237·17]
7	79[11236·15]	197[11127·15]	27487[11236·14]	30317[11237·15]
8	73[11138·19]	30211[11247·19]	13043[11137·18]	431099[11248·19]
9	59[101239·19]	168713[111249·19]	209509[111238·18]	17177[11248·19]
10	5611[1111249·21]	310841[1111249·21]	2197[1101244·20]	5513[11249·21]
11	8011[111234·10·20]	127[1111336·20]	63653[1111249·19]	1019[1134·10·20]
12	1597[1111224·10·23]	4013[111134·10·23]	1873[1111248·22]	991[1234·10·23]
13	47[1111234·10·23]	109[111135·10·23]	101[111124·10·22]	31[1235·10·23]
14	[1111234·11·25]	[111135·10·25]	[111124·10·24]	[1235·11·25]

Order	16·84	$16\cdot4^3$	12^24	$12\cdot8^2$
0	[113·11]	[112·10]	[249]	[14·10]
1	[13·11]	[12·10]	7[249]	7[14·10]
2	29[14·13]	211[113·12]	211[25·11]	73[14·12]
3	37[113·12]	659[123·11]	119[25·10]	667[25·11]
4	52139[225·15]	3671[123·14]	3821[125·13]	22697[125·14]
5	9893[216·15]	2699[115·14]	18667[136·13]	4679[116·14]
6	299297[1136·17]	29593[1115·16]	490283[1137·15]	3257831[1137·16]
7	2511043[11237·15]	106361[11215·14]	193[1133·13]	22177[1127·14]
8	360131[11235·19]	624511[11135·18]	441773[1238·17]	374281[1236·18]
9	173497[11237·19]	884393[11236·18]	24799[1238·17]	204907[1238·18]
10	589[11208·21]	14113[10236·20]	2647[1148·19]	4783[1138·20]
11	29437[11348·20]	2771[11138·19]	677[1249·18]	2251[1248·19]
12	3307[11348·23]	2579[11238·22]	2707[224·10·21]	1277[1339·22]
13	[10249·23]	23[11238·22]	23[224·10·21]	61[1349·22]
14	[11349·25]	[11239·24]	[224·11·23]	[1349·24]

Order	$12\cdot84^2$	$12\cdot4^4$	8^34	8^24^3
0	[139]	[128]	[3·10]	[29]
1	7[139]	7[128]	7[3·10]	7[29]
2	227[14·11]	47[3·10]	235[4·12]	9[·11]
3	257[23·10]	35[39]	281[13·11]	947[13·10]
4	97523[125·13]	319[110·12]	4177[112·14]	1811[110·13]
5	*245[5·13]*	*69971[124·12]*	643[111·14]	27737[113·13]
6	61219[1106·15]	3151259[1125·14]	98797[124·16]	1783141[125·15]
7	9479[1026·13]	34637[1115·12]	907[114·14]	20627[115·13]
8	12738433[1237·17]	1202651[1126·16]	37151[215·18]	1772417[225·17]
9	46223[1136·17]	115769[1126·16]	228503[236·18]	547889[226·17]
10	53593[238·19]	9841[1125·18]	50333[236·20]	151331[226·19]
11	1937[1228·18]	1823[1127·17]	391[137·19]	127[223·18]
12	1571[1238·21]	157[1028·20]	1163[336·22]	2329[227·21]
13	53[1239·21]	[1117·20]	[325·22]	37[227·21]
14	[1239·23]	[1129·22]	[337·24]	[227·23]

Order	84^6	4^7
0	[18]	[7]
1	7[18]	7[7]
2	251[2·10]	259[19]
3	1051[12·9]	77[8]
4	182843[113·12]	2107[1·11]
5	24391[103·12]	1603[1·11]
6	127741[104·14]	29771[3·13]
7	1313[4·12]	2609[3·11]
8	423697[115·16]	29771[4·15]
9	54827[115·16]	1603[3·15]
10	11455[106·18]	2107[4·17]
11	47[6·17]	77[4·16]
12	95[106·20]	259[6·19]
13	29[117·20]	7[6·19]
14	[117·22]	[7·21]

REFERENCES

CRAIG, C. C. (1928). *Metron*, **7**, 3.
FISHER, R. A. (1929). *Proc. Lond. Math. Soc.* (2), **30**, 199.
FISHER, R. A. (1930). *Proc. Roy. Soc.* A, **130**, 16.
FISHER, R. A. & WISHART, J. (1931). *Proc. Lond. Math. Soc.* (2), **33**, 195.
GEARY, R. C. (1933). *Biometrika*, **25**, 184.
HSU, C. T. & LAWLEY, D. N. (1940). *Biometrika*, **31**, 238.
McKAY, A. T. (1933). *Biometrika*, **25**, 411.
PEPPER, J. (1932). *Biometrika*, **24**, 55.
WISHART, J. (1930). *Biometrika*, **22**, 224.

THE ASYMPTOTICAL DISTRIBUTION OF RANGE IN SAMPLES FROM A NORMAL POPULATION

By G. ELFVING, *Helsingfors*

1. *Introductory.* Consider a sample of n observations, taken from an infinite normal population with the mean 0 and the standard deviation 1. Let $\mathbf{a}$ be the smallest and $\mathbf{b}$ the greatest of the observed values. Then $\mathbf{w} = \mathbf{b} - \mathbf{a}$ is the *range* of the sample.

For certain statistical purposes knowledge of the sampling distribution of range is needed. The distribution function, however, involves a rather complicated integral, whose exact calculation is, for $n > 2$, impossible. Tippett (1925), E. S. Pearson (1926, 1932) and McKay & Pearson (1933) have studied and calculated the mean, the standard deviation and the Pearson constants β_1, β_2 of the range. Fitting appropriate Pearson curves to the distribution by means of these parameters, Pearson (1932) has computed approximate percentage points for it. Later on, Hartley (1942) and Hartley & Pearson (1942) have, by numerical integration, tabulated the distribution function for $n = 2, \ldots, 20$.

As pointed out by Pearson, the distribution of range is very sensitive to departures from normality in the tails of the parental distribution. The effect of such departures becoming more perceptible for increasing n, the practical importance of the range distribution is, perhaps, small for large samples. Nevertheless, it seems to be at least of theoretical interest to investigate the *asymptotical* distribution of range for $n \to \infty$. This is the purpose of the present paper.* The results are summarized in a theorem at the end of the inquiry.

2. *The exact distribution. Transformations.* The joint-frequency function of the extremes $\mathbf{a}$, $\mathbf{b}$ reads, as well known,

$$f_{\mathbf{ab}}(a,b) = n(n-1)\,\phi(a)\,\phi(b)\,[\Phi(b)-\Phi(a)]^{n-2}\dagger \tag{2·1}$$

(cf. e.g. Cramér, 1945, p. 370). Let $\mathbf{u} = \frac{1}{2}(\mathbf{a}+\mathbf{b})$ denote the arithmetical mean of the extreme values of the sample. Making in (2·1) the transformation $\mathbf{a} = \mathbf{u} - \frac{1}{2}\mathbf{w}$, $\mathbf{b} = \mathbf{u} + \frac{1}{2}\mathbf{w}$ and integrating with respect to u, we find for the frequency function of the range the expression

$$f_{\mathbf{w}}(w) = n(n-1)\int_{-\infty}^{\infty} \phi(u-\tfrac{1}{2}w)\,\phi(u+\tfrac{1}{2}w)\,[\Phi(u+\tfrac{1}{2}w)-\Phi(u-\tfrac{1}{2}w)]^{n-2}\,du. \tag{2·2}$$

The object of our inquiry is the limiting form of the distribution (2·2). It proves, however, more advantageous to pass to the limit in the joint distribution of $\mathbf{a}$, $\mathbf{b}$ or $\mathbf{u}$, $\mathbf{w}$, before integrating with respect to u.

The asymptotical distribution of $\mathbf{a}$ and $\mathbf{b}$ has been investigated by Fisher & Tippett (1928), and Gumbel (1936) (cf. also Cramér, 1945, p. 376). According to these authors, we have

$$\left.\begin{aligned} E(\mathbf{u}) &= 0, & D(\mathbf{u}) &= O(\log^{-\frac{1}{2}} n),\\ E(\mathbf{w}) &= 2\sqrt{(2\log n)} + O\left(\frac{\log\log n}{\sqrt{(\log n)}}\right), & D(\mathbf{w}) &= O(\log^{-\frac{1}{2}} n).\end{aligned}\right\} \tag{2·3}$$

From the formulae quoted it is seen that $\mathbf{u} \to 0$, $\mathbf{w} \to \infty$ *in probability* as $n \to \infty$. Our first task must, consequently, be a transformation of the variables $\mathbf{a}$, $\mathbf{b}$—or $\mathbf{u}$, $\mathbf{w}$—depending on n and intended to stabilize the probability mass, in order to provide a limiting distribution.

* Prof. H. Wold has kindly directed my attention to this problem.

† $\Phi(x)$ denotes the distribution function and $\phi(x) = \Phi'(x)$ the frequency function of the normal distribution with mean at $x = 0$ and unit standard deviation.

Following the example of the authors mentioned above, we should have to introduce the new variables

$$\mathbf{a}' = n\Phi(\mathbf{a}), \quad \mathbf{b}' = n\Phi(-\mathbf{b}).$$

For our purpose it proves, however, advantageous to subject $\mathbf{a}'$ and $\mathbf{b}'$ to a new transformation, independent of n, taking

$$\left.\begin{aligned} \mathbf{x}e^{\mathbf{y}} &= 2n\Phi(\mathbf{a}) \quad = 2n\Phi(-\tfrac{1}{2}\mathbf{w}+\mathbf{u}), \\ \mathbf{x}e^{-\mathbf{y}} &= 2n\Phi(-\mathbf{b}) = 2n\Phi(-\tfrac{1}{2}\mathbf{w}-\mathbf{u}). \end{aligned}\right\} \tag{2·4}$$

Conversely,

$$\left.\begin{aligned} \mathbf{x} &= 2n\sqrt{[\Phi(\mathbf{a})\,\Phi(-\mathbf{b})]} = 2n\sqrt{[\Phi(-\tfrac{1}{2}\mathbf{w}+\mathbf{u})\,\Phi(-\tfrac{1}{2}\mathbf{w}-\mathbf{u})]}, \\ \mathbf{y} &= \tfrac{1}{2}\log\frac{\Phi(\mathbf{a})}{\Phi(-\mathbf{b})} = \tfrac{1}{2}\log\frac{\Phi(-\tfrac{1}{2}\mathbf{w}+\mathbf{u})}{\Phi(-\tfrac{1}{2}\mathbf{w}-\mathbf{u})}. \end{aligned}\right\} \tag{2·5}$$

As $\mathbf{a}\leqq\mathbf{b}$ and thus $\Phi(\mathbf{a})+\Phi(-\mathbf{b})\leqq 1$, it follows from (2·4), that $\mathbf{x}$, $\mathbf{y}$ are subjected to the restrictions

$$\mathbf{x}\geqq 0, \quad \mathbf{x}\cosh\mathbf{y}\leqq n. \tag{2·6}$$

Performing the transformation, we find

$$\left|\frac{\partial(a,b)}{\partial(x,y)}\right| = \frac{x}{2n^2\phi(a)\,\phi(b)}, \tag{2·7}$$

and thus, letting $f_n(x,y)$ denote the joint-frequency function of $\mathbf{x}$, $\mathbf{y}$,

$$f_n(x,y) = \frac{n-1}{2n}x\left(1-\frac{x\cosh y}{n}\right)^{n-2}. \tag{2·8}$$

This formula is valid in the region (2·6); outside of it, we have to put $f_n(x,y) = 0$.

The new variables $\mathbf{x}$, $\mathbf{y}$ depend, of course, on $\mathbf{u}$ as well as $\mathbf{w}$. It will, however, be shown later, that $\mathbf{x}$, for large n, tends to coincide with the variable

$$\mathbf{x}^* = 2n\Phi(-\tfrac{1}{2}\mathbf{w}),$$

which depends exclusively on $\mathbf{w}$. *For testing purposes, the former variable may thus, in* large samples, be used as a substitute for the range. These considerations justify the transformation (2·4) as well as a closer study of the distribution of $\mathbf{x}$ and its limiting form.

3. *Limit passage and remainder term.* The limiting form of the joint-frequency function (2·8) is immediately seen to be

$$f(x,y) = \tfrac{1}{2}xe^{-x\cosh y} \quad (x\geqq 0). \tag{3·1}$$

The integral of this function, taken over the whole half-plane $x\geqq 0$, is easily seen to equal 1; (3·1) is, consequently, the frequency function of a well-determined two-dimensional distribution.

Let the marginal distribution functions in x, corresponding to (2·8) and (3·1), be denoted by $F_n(x)$ and $F(x)$ respectively. Our next task will be to estimate the remainder $|F_n(x)-F(x)|$, which is, obviously, at most equal to the integral

$$\Delta_n = \int_0^x\int_0^\infty 2\,|f_n(\xi,\eta)-f(\xi,\eta)|\,d\xi\,d\eta. \tag{3·2}$$

To begin with, we estimate the quotient f_n/f upwards. By differentiation with respect to the variable $z = x\cosh y$, this quotient is found to attain the maximum value

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)^{n-2}e^2 = 1+\frac{1}{n}+O\left(\frac{1}{n^2}\right)$$

for $z = 2$. We thus find, for example,

$$\frac{f_n}{f} < 1+\frac{3}{2n} \quad (n\geqq 5). \tag{3·3}$$

For the further estimations, it proves necessary to divide the domain of integration in (3·2) into an interior and an exterior part by means of a convenient abscissa $\eta = y$. In order to secure the Maclaurin expansion of $\log\left(1-\frac{1}{n}\xi\cosh\eta\right)$ within the interior region, we have to choose y so as to satisfy the inequality $\frac{x\cosh y}{n} \leqq k$ with an appropriate $k<1$. Taking, for simplicity, $k = 1-\sqrt{\tfrac{1}{2}}$ and observing that $\cosh y \leqq e^y$, we see that the condition mentioned is fulfilled if

$$e^y \leqq \frac{n}{x}(1-\sqrt{\tfrac{1}{2}}). \tag{3·4}$$

Now we may estimate f_n/f downwards in the interior domain of integration. Expanding $\log\left(1-\frac{1}{n}\xi\cosh\eta\right)$, we find

$$\log\frac{f_n}{f} = \log\left(1-\frac{1}{n}\right)+\frac{2}{n}\xi\cosh\eta-\frac{n-2}{2n^2}\xi^2\cosh^2\eta\left(1-\vartheta\frac{\xi\cosh\eta}{n}\right)^{-2} \quad (0<\vartheta<1). \tag{3·5}$$

According to the determination of y, the remainder factor is seen to be <2 for $\xi\leqq x$, $\eta\leqq y$. For $n\geqq 3$, we have $\log\left(1-\frac{1}{n}\right)>-\frac{3}{2n}$. Omitting, further, the positive term in (3·5) and replacing $n-2$ by n, we find

$$\frac{f_n(\xi,\eta)}{f(\xi,\eta)}-1>\log\frac{f_n(\xi,\eta)}{f(\xi,\eta)}>-\frac{\frac{3}{2}+\xi^2\cosh^2\eta}{n};$$

hence, combining with (3·3),

$$\left|\frac{f_n(\xi,\eta)}{f(\xi,\eta)}-1\right|<\frac{\frac{3}{2}+\xi^2\cosh^2\eta}{n} \quad (\xi\leqq x, \eta\leqq y;\ n\geqq 5). \tag{3·6}$$

In the exterior domain of integration, (3·3) directly yields

$$|f_n(\xi,\eta)-f(\xi,\eta)|<f(\xi,\eta) \quad (\xi\leqq x, \eta\geqq y). \tag{3·7}$$

We proceed to the estimation of the integral (3·2), denoting its interior and exterior part by I_1 and I_2 respectively. For the former we have, according to (3·6), the inequality

$$I_1 = \int_0^x\int_0^y\left|\frac{f_n}{f}-1\right|2f\,d\xi\,d\eta<\frac{1}{n}\int_0^x\int_0^y(\tfrac{3}{2}\xi+\xi^3\cosh^2\eta)\,e^{-\xi\cosh\eta}\,d\xi\,d\eta, \tag{3·8}$$

for the latter, according to (3·7),

$$I_2 = \int_0^x\int_y^\infty 2\,|f_n-f|\,d\xi\,d\eta<\int_0^x\int_y^\infty \xi e^{-\xi\cosh\eta}\,d\xi\,d\eta. \tag{3·9}$$

The integration with respect to ξ may be explicitly performed. We have, in fact, putting for brevity $\cosh\eta = a$,

$$\int_0^x \xi e^{-a\xi}\,d\xi = \frac{1}{a^2}\{1-e^{-ax}[1+ax]\}, \tag{3·10}$$

$$\int_0^x \xi^3 e^{-a\xi}\,d\xi = \frac{6}{a^4}\left\{1-e^{-ax}\left[1+ax+\frac{(ax)^2}{2}+\frac{(ax)^3}{6}\right]\right\}. \tag{3·11}$$

In order to deduce remainder formulas for (a) moderate, (b) small x, we omit in (3·10) and (3·11), (a) all the negative terms, (b) the terms with x^2 and x^3. According to the Maclaurin expansion

$$e^{ax} = 1+ax+e^{\vartheta ax}\frac{(ax)^2}{2} \quad (0<\vartheta<1),$$

the expression in curled brackets in (3·10) is at most equal to $\frac{1}{2}a^2x^2$. Inserting these estimations in (3·8), we obtain for the interior integral the inequalities

$$I_1 < \frac{15}{2n}\int_0^y \frac{d\eta}{\cosh^2\eta} = \frac{15}{2n}\operatorname{tgh} y < \frac{15}{2n}, \tag{3·12a}$$

$$I_1 < \frac{15}{4n}x^2\int_0^y d\eta \leqslant \frac{4x^2}{n}y. \tag{3·12b}$$

For the exterior integral, (3·10) yields

$$I_2 < \int_y^\infty \frac{d\eta}{\cosh^2\eta} = 1 - \operatorname{tgh} y < 2e^{-2y}. \tag{3·13}$$

Finally, we have to join the results (3·12) and (3·13). Combining, first, (3·12a) with (3·13) and determining e^{-y} from (3·4) (taken with the equality sign), we obtain, after some slight simplifications in the numerical coefficients,

$$\Delta_n < \frac{8}{n}\left(1 + \frac{3x^2}{n}\right) \quad (n \geqq 5). \tag{3·14a}$$

Combining, on the other hand, (3·12b) with (3·13), we find

$$\Delta_n < \frac{4x^2}{n}y + 2e^{-2y}.$$

This expression attains, for fixed x and n, its minimum when $y = \log\frac{\sqrt{n}}{x}$. For $n \geqq 12$, this value of y also satisfies (3·4), and we obtain, as a parallel estimate to (3·14a),

$$\Delta_n < \frac{4x^2}{n}\left(\log\frac{\sqrt{n}}{x} + \tfrac{1}{2}\right) \quad (n \geqq 12). \tag{3·14b}$$

The formulas (3·14a, b) are both valid for all positive x and all $n \geqq 12$.

4. *The asymptotical distribution.* Having established the limiting distribution of the variable x defined in (2·5), we are going to examine its properties.

The frequency function of the distribution considered reads, according to (3·1),

$$f(x) = x\int_0^\infty e^{-x\cosh y}\,dy = x\int_1^\infty \frac{e^{-xt}}{\sqrt{(t^2-1)}}\,dt. \tag{4·1}$$

Changing the order of integration, we easily find the distribution function, the mean and the variance of (4·1) to be

$$F(x) = 1 - \int_0^\infty \frac{1 + x\cosh y}{\cosh^2 y}e^{-x\cosh y}\,dy = 1 - \int_1^\infty \frac{1+xt}{t^2\sqrt{(t^2-1)}}e^{-xt}\,dt, \tag{4·1$'$}$$

$$E(\mathrm{x}) = \tfrac{1}{2}\pi, \quad D^2(\mathrm{x}) = 4 - \tfrac{1}{4}\pi^2. \tag{4·2}$$

The numerical evaluation of the distribution is much simplified by the fact that $f(x)$ as well as $F(x)$ is closely connected with certain *Bessel functions*. Denote

$$\phi(x) = \int_0^\infty e^{-x\cosh y}\,dy = \int_1^\infty \frac{e^{-xt}}{\sqrt{(t^2-1)}}\,dt. \tag{4·3}$$

By differentiation and partial integration, this function is found to satisfy the differential equation

$$\phi''(x) + \frac{1}{x}\phi'(x) - \phi(x) = 0. \tag{4·4}$$

Changing x into $-ix$, we obtain for the function $\psi(x) = \phi(-ix)$ the equation

$$\psi''(x) + \frac{1}{x}\psi'(x) + \psi(x) = 0; \tag{4·4'}$$

hence, $\psi(x)$ is a Bessel function of order zero.

In order to specify this function, we will deduce an asymptotical expression for the function (4·3), valid for large x. For this purpose, we make in the latter integral (4·3) the substitution $t = 1 + u/x$ and write

$$\left(1+\frac{u}{2x}\right)^{-\frac{1}{2}} = 1 - \vartheta\frac{u}{4x} \quad (0<\vartheta<1).$$

Performing the integration, we obtain

$$\phi(x) = \sqrt{\left(\frac{\pi}{2}\right)}\, x^{-\frac{1}{2}} e^{-x}\left(1 + O\left(\frac{1}{x}\right)\right), \tag{4·5}$$

which shows that the Bessel function $\psi(x) = \phi(-ix)$ tends to zero for $x \to +i\infty$. This function is, consequently, proportional to the *Hankel function* $H_0^{(1)}(x)$ (cf. Jahnke-Emde, 1909, p. 94). Comparing the asymptotical expressions of $\phi(x)$ and $iH_0^{(1)}(ix)$, we find the proportional factor to be $\frac{1}{2}\pi$, whence

$$f(x) = x\frac{\pi i}{2} H_0^{(1)}(ix). \tag{4·6}$$

We proceed to the calculation of $F(x)$. Every integral of $xH_0^{(1)}(x)$ is (cf. Jahnke-Emde, p. 165) of the form $xH_1^{(1)}(x) + \text{Const.}$, where $H_1^{(1)}(x)$ is the *first* order Hankel function corresponding to $H_0^{(1)}(x)$; consequently,

$$F(x) = \frac{\pi x}{2} H_1^{(1)}(ix) + C.$$

Now $\frac{\pi x}{2} H_1^{(1)}(ix)$ tends to zero as $-(\frac{1}{2}\pi x)^{\frac{1}{2}} e^{-x}$ for $x \to \infty$ (cf. Jahnke-Emde, 1909, p. 101); hence $C = 1$ and

$$F(x) = 1 - x\left[-\frac{\pi}{2} H_1^{(1)}(ix)\right]. \tag{4·7}$$

For small x, $F(x)$ has the expansion

$$F(x) = \left(\log\frac{2}{\gamma x} + \frac{1}{2}\right)\frac{x^2}{2} + \left(\log\frac{2}{\gamma x} + \frac{5}{4}\right)\frac{x^4}{16} + \dots, \tag{4·8}$$

where

$$\log\frac{2}{\gamma} = 0{\cdot}11593\dots. \tag{4·9}$$

The factors of x in (4·6) and (4·7) are tabulated in Jahnke-Emde (1909, pp. 135–6). Below, we give a short table of $f(x)$ and $F(x)$. The corresponding curves are seen in Fig. 1.

5. *Connexion between the variable* **x** *and the range.* We now turn back to the original object of our inquiry: the asymptotical distribution of the range.

Consider the variable

$$\mathbf{x} = 2n\sqrt{[\Phi(-\tfrac{1}{2}\mathbf{w}+\mathbf{u})\,\Phi(-\tfrac{1}{2}\mathbf{w}-\mathbf{u})]} \tag{5·1}$$

introduced in (2·4). As mentioned earlier,

$$\mathbf{w} \to \infty, \quad \mathbf{u} \to 0 \text{ in probability} \quad (n \to \infty). \tag{5·2}$$

Under such circumstances, for large n, **x** may be expected to behave substantially as the variable

$$\mathbf{x}^* = 2n\Phi(-\tfrac{1}{2}\mathbf{w}), \tag{5·3}$$

which depends *exclusively on the range.*

We shall now prove that $\mathbf{x}^*/\mathbf{x} \to 1$ in probability as $n \to \infty$. According to the well-known asymptotic formula

$$\Phi(-x) = \frac{1}{x\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\left(1 - \frac{\vartheta}{x^2}\right) \quad (x > 0);\ \ 0\,(<\vartheta < 1),$$

we may, for $|\mathbf{u}| < \frac{1}{2}\mathbf{w}$, write

$$\frac{\mathbf{x}^*}{\mathbf{x}} = e^{\frac{1}{2}\mathbf{u}^2}\left(1 - \frac{4\mathbf{u}^2}{\mathbf{w}^2}\right)^{\frac{1}{2}} \{1 + O[(\tfrac{1}{2}\mathbf{w} - |\mathbf{u}|)^{-2}]\}.$$

x	$f(x)$	$F(x)$	x	$f(x)$	$F(x)$
0·0	0·0000	0·0000	1·5	0·3207	0·5839
0·1	0·2427	0·0146	2·0	0·2278	0·7202
0·2	0·3505	0·0448	2·5	0·1559	0·8153
0·3	0·4118	0·0832	3·0	0·1042	0·8795
0·4	0·4458	0·1262	4·0	0·0446	0·9501
0·5	0·4622	0·1718	5·0	0·0185	0·9798
0·6	0·4665	0·2183	6·0	0·0075	0·9919
0·7	0·4624	0·2648	7·0	0·0030	0·9968
0·8	0·4522	0·3106	8·0	0·0012	0·9988
0·9	0·4380	0·3552	9·0	0·0005	0·9995
1·0	0·4210	0·3981	10·0	0·0002	0·9998

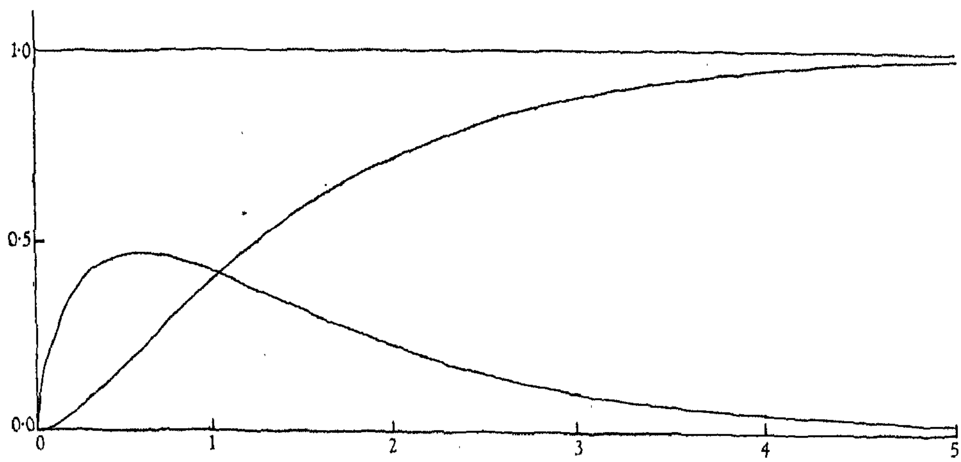


Fig. 1

Given an arbitrary $\epsilon > 0$, we obviously may find two positive numbers u_ϵ and w_ϵ $(> u_\epsilon)$ such that

$$\left|\frac{\mathbf{x}^*}{\mathbf{x}} - 1\right| < \epsilon \quad \text{if} \quad \mathbf{w} \geqq w_\epsilon, \quad |\mathbf{u}| \leqq u_\epsilon. \tag{5·4}$$

On account of (5·2), we may, on the other hand, choose n_ϵ so that the probability of the simultaneous validity of the latter inequalities in (5·4) exceeds $1 - \epsilon$ if $n \geqq n_\epsilon$. Consequently,

$$P\left\{\left|\frac{\mathbf{x}^*}{\mathbf{x}} - 1\right| < \epsilon\right\} > 1 - \epsilon \quad (n \geqq n_\epsilon), \tag{5·5}$$

which proves our statement.

As shown in section 3, the distribution function $F_n(x)$ of $\mathbf{x}$ converges to $F(x)$ as $n \to \infty$. Since $F(0) = 0$, it follows from (5·5), by a well-known method of argument, that the distribution function $F_n^*(x)$ of $\mathbf{x}^*$ converges to the same limiting function. The asymptotical distribution of the range, suitably transformed, is hereby established.

For practical purposes, it would, of course, be desirable to possess a reasonably accurate estimate of the remainder $F_n^*(x) - F(x)$, or at least an estimate of the difference $F_n^*(x) - F_n(x)$, to be combined with the results (3·14).

For $n = 20$, the accuracy of $F(x)$ as substitute for $F_n^*(x)$ may be checked by means of Hartley's (1942) tables. The discrepancy amounts to about 0·004 for $x = 0{\cdot}1$, 0·025 for $x = 1$ and 0·010 for $x = 4$.

The theoretical evaluation of $F_n^*(x) - F(x)$ seems to be somewhat complicated and, besides, of little use since $\mathbf{x}^*$, for most purposes, may be replaced by $\mathbf{x}$. A few remarks concerning the relations between $\mathbf{x}$, $\mathbf{x}^*$ and their distribution functions will, however, be added below.

To begin with, we note that always $\mathbf{x} \leqq \mathbf{x}^*$, the equality sign being valid only if $\mathbf{u} = 0$. Consider, in fact, the function $x(u)$, defined by (5·1) for a fixed w. Inserting for Φ its analytical expression, we easily find that $D^{(2)} \log x(u) \leqq 0$ for all u. Hence, $x(u)$ has no minimum and at most one maximum, and the latter is, by symmetry, seen to be attained for $u = 0$, being thus equal to $\mathbf{x}^*$.

From $\mathbf{x} \leqq \mathbf{x}^*$, it follows that $F_n^*(x) \leqq F_n(x)$ for all x. We will show that the difference $F_n(x) - F_n^*(x)$ may be expressed as a double integral.

The variables $\mathbf{u}$ and $\mathbf{w}$ are, according to (2·4), well-determined functions of $\mathbf{x}$ and $\mathbf{y}$ in the region (2·6); and so is the variable $\mathbf{x}^*$, on account of (5·3).

On the *level curve* $\mathbf{x}^* = x_0$, $\mathbf{w}$ has a constant value w_0, determined by

$$2n\Phi(-\tfrac{1}{2}w_0) = x_0, \tag{5·6}$$

and this curve is, consequently, given in parametric form by the equations

$$x = 2n\sqrt{[\Phi(-\tfrac{1}{2}w_0+u)\,\Phi(-\tfrac{1}{2}w_0-u)]}, \quad y = \tfrac{1}{2}\log\frac{\Phi(-\tfrac{1}{2}w_0+u)}{\Phi(-\tfrac{1}{2}w_0-u)}, \tag{5·7}$$

where u runs through all values from $-\infty$ to $+\infty$. The latter function (5·7) being, obviously, monotonously increasing, we may imagine u eliminated, writing (5·7) in the form

$$x = \xi_n(x_0, y) \quad (-\infty < y < \infty). \tag{5·7'}$$

From the proof of the inequality $\mathbf{x} \leqq \mathbf{x}^*$ given above, it follows that the function (5·7′) has a single maximum for $y = 0$. When $y \to \pm\infty$, the function obviously tends to zero.

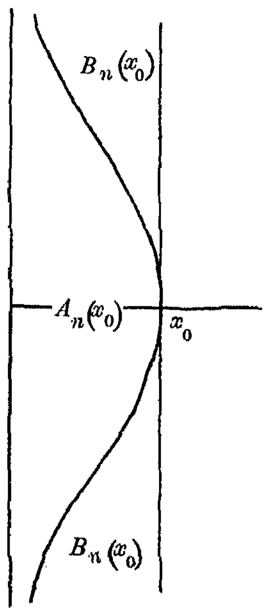


Fig. 2

The inequality $\mathbf{x}^* \leqq x_0$ is fulfilled on the left side of the curve (5·7′), the inequality $\mathbf{x} \leqq x_0$ on the left side of the straight line $\mathbf{x} = x_0$. Let us for brevity denote the regions (cf. fig. 2)

$$0 \leqq \mathbf{x} \leqq \xi_n(x_0, \mathbf{y}), \quad \xi_n(x_0, \mathbf{y}) < \mathbf{x} \leqq x_0 \tag{5·8}$$

by $A_n(x_0)$ and $B_n(x_0)$ respectively. The difference $F_n(x_0) - F_n^*(x_0)$ is, then, the probability of the points $\mathbf{x}$, $\mathbf{y}$ falling within the region $B_n(x_0)$. Dropping the indices 0, we thus obtain the expression sought for

$$F_n(x) - F_n^*(x) = \iint_{B_n(x)} f_n(\xi, \eta)\, d\xi\, d\eta. \tag{5·9}$$

Comparing, finally, the transformed range distribution function $F_n^*(x)$ directly with its limiting form $F(x)$, we find

$$\begin{aligned} F_n^*(x) - F(x) &= [F_n(x) - F(x)] - [F_n(x) - F_n^*(x)] \\ &= \iint_{\xi \leqq x} (f_n - f)\, d\xi\, d\eta - \iint_{B_n(x)} f_n\, d\xi\, d\eta \\ &= \iint_{A_n(x)} (f_n - f)\, d\xi\, d\eta - \iint_{B_n(x)} f\, d\xi\, d\eta. \end{aligned} \tag{5·10}$$

The former integral is, obviously, at most equal to the remainder expression Δ_n in (3·2), estimated in (3·14).

6. *Conclusion.* Our main results may be summarized in the following theorem:

THEOREM. Consider a sample of n observations from an infinite normal population with mean 0 and standard deviation 1. Let $\mathbf{a}$ be the smallest, $\mathbf{b}$ the greatest of the observed values, and put

$$\mathbf{x} = 2n\sqrt{[\Phi(\mathbf{a})\,\Phi(-\mathbf{b})]}, \quad \mathbf{x}^* = 2n\Phi\left(-\frac{\mathbf{b}-\mathbf{a}}{2}\right),$$

the latter variable being evidently a simple transformation of the range of the sample. Then

(1) $\mathbf{x} \leqq \mathbf{x}^*$; $\mathbf{x}^*/\mathbf{x} \to 1$ in probability $(n \to \infty)$.

(2) *The distribution functions $F_n(x)$ and $F_n^*(x)$ of $\mathbf{x}$ and $\mathbf{x}^*$ tend, for $n \to \infty$, to the* common limit

$$F(x) = 1 - \int_1^\infty \frac{1+xt}{t^2\sqrt{(t^2-1)}}\, e^{-xt}\, dt = 1 + \frac{\pi x}{2} H_1^{(1)}(ix),$$

where $H_1^{(1)}(z)$ is the first order Bessel function, which vanishes as $-\left(\frac{\pi z}{2i}\right)^{-\frac{1}{2}} e^{iz}$ for $z \to +i\infty$.

(3) For $n \geqq 12$, $F_n(x)$ satisfies the inequalities

$$|F_n(x) - F(x)| < \frac{8}{n}\left(1 + \frac{3x^2}{n}\right), \quad |F_n(x) - F(x)| < \frac{4x^2}{n}\left(\log\frac{\sqrt{n}}{x} + \frac{1}{2}\right).$$

7. *Generalization.* A great part of our conclusions does not presuppose the normality of the parental population. Thus, the distribution (2·8) of the variables $\mathbf{x}$, $\mathbf{y}$ defined by (2·5) is the same for any continuous probability law and so; consequently, is its limiting form; however, if the parental distribution is non-symmetrical, with distribution function $G(x)$, say, the factor $\Phi(-\mathbf{b})$ in (2·5) must, of course, be replaced by $1 - G(\mathbf{b})$ instead of $G(-\mathbf{b})$, and the variable $\mathbf{x}^*$ is to be defined by

$$\mathbf{x}^* = 2n\sqrt{\{G(-\tfrac{1}{2}\mathbf{w})\,[1 - G(\tfrac{1}{2}\mathbf{w})]\}}.$$

The proof of the statement $\mathbf{x}^*/\mathbf{x} \to 1$ requires, however, convenient assumptions concerning the parental distribution. It can be proved that the assertion mentioned—and, consequently, the theorem stated above—are valid if the frequency function of this distribution is of the form

$$g(x) = C\exp\left[-\frac{1}{p}|x|^p\right],$$

where $1 < p \leqq 2$.

REFERENCES

CRAMÉR, H. (1945). *Mathematical Methods of Statistics*. Uppsala.

FISHER, R. A. & TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24**, 180.

GUMBEL, E. J. (1936). Les valeurs extrêmes des distribution statistiques. *Ann. Inst. Poincaré*, **5**, 115–58.

HARTLEY, H. O. (1942). The range in random samples. *Biometrika*, **32**, 334–48.

HARTLEY, H. O. & PEARSON, E. S. (1942). The probability integral of the range in samples of n observations from a normal population. *Biometrika*, **32**, 301–10.

JAHNKE, E. & EMDE, F. (1909). *Funktionentafeln*. Leipzig and Berlin.

MCKAY, A. T. & PEARSON, E. S. (1933). A note on the distribution of range in samples of n. *Biometrika*, **25**, 415–20.

PEARSON, E. S. (1926). A further note on the distribution of range in samples, taken from a normal population. *Biometrika*, **18**, 173–94.

PEARSON, E. S. (1932). The percentage limits for the distribution of range in samples from a normal population. *Biometrika*, **24**, 404–17.

TIPPETT, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika*, **17**, 364–87.

LIMITS OF THE RATIO OF MEAN RANGE TO STANDARD DEVIATION*

By R. L. PLACKETT, B.A.

The ratio of mean range $\bar{w}_n$ in samples of n to population standard deviation σ, which has been denoted by d_n, is used in control chart work (when the population is assumed normal) to estimate σ from the ranges of a set of small samples. On comparing the series of values of d_n for different n when the parent population is rectangular with the series when it is normal (see table below), it is clear that for $n \leqslant 12$ the two series agree to within less than 10 %. With this in mind, the question arises: what are the limiting values of d_n for a given n? It is shown here that populations exist for which d_n is arbitrarily near to zero, while for no population will d_n exceed the value

$$n\sqrt{\left(\frac{2}{(2n-1)!}\{(2n-2)!-[(n-1)!]^2\}\right)}.$$

We consider a population whose distribution function is $F(x)$ and which extends from $-a$ to $+a$ so that $F(-a)=0$ and $F(a)=1$. The population in the first place may have any finite limits, but there is no loss in generality in supposing these. It is required to find limits to the ratio

$$d_n=\frac{\int_{-a}^{a}[1-F^n-(1-F)^n]\,dx}{\left[\int_{-a}^{a}x^2dF-\left(\int_{-a}^{a}x\,dF\right)^2\right]^{\frac{1}{2}}}. \tag{1}$$

We apply the calculus of variations and find the extremes of d_n in the class of functions F such that $F(-a)=0$ and $F(a)=1$; the case is thus one of fixed end-points. Suppose that $F(x)=u(x)$ gives an extreme value and form the functions $F(x)=u(x)+tv(x)$; for t suitably near to zero, all these will be permissible distribution functions, i.e. monotonically increasing, provided $v(-a)=v(a)=0$. Then for $t=0$, $d/dt(d_n)$ is zero for all functions $v(x)$.

Since
$$d_n=\frac{\int_{-a}^{a}[1-(u+tv)^n-(1-u-tv)^n]\,dx}{\left[\int_{-a}^{a}x^2(u'+tv')\,dx-\left(\int_{-a}^{a}x(u'+tv')\,dx\right)^2\right]^{\frac{1}{2}}},$$

$$\left[\frac{d}{dt}(d_n)\right]_{t=0}=\frac{\begin{bmatrix}2n\left[\int_{-a}^{a}x^2u'dx-\left(\int_{-a}^{a}xu'dx\right)^2\right]\left[\int_{-a}^{a}(1-u)^{n-1}v\,dx-\int_{-a}^{a}u^{n-1}v\,dx\right]\\ -\left[\int_{-a}^{a}(1-u^n-(1-u)^n)\,dx\right]\left[\int_{-a}^{a}x^2v'dx-2\left(\int_{-a}^{a}xu'dx\right)\left(\int_{-a}^{a}xv'dx\right)\right]\end{bmatrix}}{2\left[\int_{-a}^{a}x^2u'dx-\left(\int_{-a}^{a}xu'dx\right)^2\right]^{\frac{3}{2}}}$$
$$=0.$$

Now $\int_{-a}^{a}x^2v'dx=-2\int_{-a}^{a}xv\,dx$ since $v(a)=v(-a)=0$, and by the same condition

$$\int_{-a}^{a}xv'dx=-\int_{-a}^{a}v\,dx.$$

* Communication from the National Physical Laboratory.

The numerator now becomes of the form $\int_{-a}^{a} s(x)\,v(x)\,dx$, and this must be zero for all functions $v(x)$; it is therefore concluded that $s(x)$ is identically equal to zero. In fact

$$n\left[\int_{-a}^{a} x^2u'\,dx - \left(\int_{-a}^{a} xu'\,dx\right)^2\right][(1-u)^{n-1}-u^{n-1}]$$
$$= \left[\int_{-a}^{a}\{1-u^n-(1-u)^n\}\,dx\right]\left[\int_{-a}^{a} xu'\,dx - x\right],$$

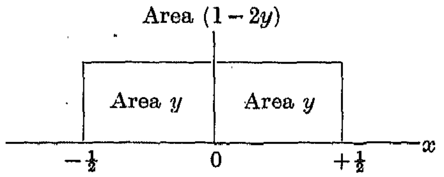
so that if μ is the mean, σ the standard deviation, $\overline{w}_n$ the mean range in samples of n, and $F(x)$ the distribution function of the population which gives an extreme value to d_n, we have

$$\overline{w}_n(x-\mu) = n\sigma^2[F^{n-1}-(1-F)^{n-1}].$$

Put $x = -a$ and obtain $n\sigma^2 = \overline{w}_n(\mu+a)$; $x = a$ gives $n\sigma^2 = \overline{w}_n(a-\mu)$ whence $\mu = 0$ and

$$x = a[F^{n-1}-(1-F)^{n-1}]. \tag{2}$$

This distribution must give an upper limit to d_n since if we consider a distribution of the type below:



$$F(x) = (2x+1)y \qquad -\tfrac{1}{2} \leqslant x < 0$$
$$F(x) = 1-(1-2x)y \qquad 0 < x \leqslant \tfrac{1}{2}$$
$$y \leqslant \tfrac{1}{2}$$

the ratio (1) for $y = O(n^{-3})$ is approximately $\sqrt{(3/2)}\,n\sqrt{y}$ which can be made as small as we please.

Reverting therefore to equation (2) we note that since $a\overline{w}_n = n\sigma^2$, $d_n(\max.) = n\sigma/a$.

$$\sigma^2 = \int_{-a}^{a} x^2\,dF - \left(\int_{-a}^{a} x\,dF\right)^2$$
$$= a^2\int_{-a}^{a}[F^{n-1}-(1-F)^{n-1}]^2\,dF.$$

Therefore
$$\sigma^2/a^2 = \frac{2}{2n-1} - 2B(n,n),$$

i.e.
$$d_n(\max.) = n\sqrt{\left(\frac{2}{(2n-1)!}\{(2n-2)!-[(n-1)!]^2\}\right)}. \tag{3}$$

It is of interest to note that all the foregoing analysis may be carried out with a equal to any finite value and so we may take the limit as $a \to \infty$, and equation (3), which is independent of a, will still hold.

It is easy to verify, by Stirling's formula or otherwise, that as n increases $[(n-1)!]^2$ becomes negligible compared with $(2n-2)!$.

Consequently, for large n,
$$d_n(\max.) \simeq n\sqrt{\{2/(2n-1)\}}$$
$$= \sqrt{\left(n+\frac{1}{2-1/n}\right)},$$

i.e.
$$d_n(\max.) \simeq \sqrt{(n+\tfrac{1}{2})}. \tag{4}$$

The probability density function of (2) is obtained by differentiation and is

$$f(x) = \frac{1}{a(n-1)[F^{n-2}+(1-F)^{n-2}]}, \tag{5}$$

so that (2) and (5) are the parametric equations of the curve in terms of its distribution function. Thus for $n > 2$, $f(0) = \dfrac{2^{n-3}}{a(n-1)}$ and $f(\pm a) = \dfrac{1}{a(n-1)}$. The distributions (2) are readily seen to be unimodal and symmetrical about $x = 0$. For $n = 2, 3$ they are rectangular. For $F > \frac{1}{2}$ and large n, $aF^{n-1} \simeq x$. Hence $F^{n-2} \simeq \dfrac{x}{a}$, $f(x) \simeq \dfrac{1}{x(n-1)}$. Similar considerations for $F < \frac{1}{2}$ show that for large n and $x \neq 0$,

$$f(x) \simeq \frac{1}{|x|(n-1)}.$$

From (4), $\sigma \sim a/\sqrt{n}$. Consequently, for any finite a, as $n \to \infty$ the distributions (2) tend to a single ordinate at $x = 0$. This should be compared with the limiting case giving $d_n \to 0$ for fixed n illustrated with the diagram above. The limiting form of the two distributions is the same but the approach to the limit with increasing n is quite different. There is no approach to normality.

Following is a table of d_n(max.) and of d_n in samples from normal and rectangular populations for $n = 2, \ldots, 12$. The quantity $\sqrt{(n + \frac{1}{2})}$ is also included to see how closely (4) is approximated. The values of d_n (normal) are obtained from the paper by E. S. Pearson (1942). For a rectangular distribution d_n is simply $2\sqrt{3}(n-1)/(n+1)$.

n	$\sqrt{(n+\frac{1}{2})}$	d_n (max.)	d_n (normal)	d_n (rectangular)
2	1·58114	1·15470	1·128	1·15470
3	1·87083	1·73205	1·693	1·73205
4	2·12132	2·08395	2·059	2·07846
5	2·34521	2·34013	2·326	2·30940
6	2·54951	2·55333	2·534	2·47436
7	2·73861	2·74414	2·704	2·59808
8	2·91548	2·92076	2·847	2·69430
9	3·08221	3·08685	2·970	2·77128
10	3·24037	3·24440	3·078	2·83426
11	3·39116	3·39466	3·173	2·88675
12	3·53553	3·53860	3·258	2·93116

Some values of d_n for a number of symmetrical populations were given by Pearson & Adyanthaya (1928) and have been reproduced with some figures for one skew population in *Tables for Statisticians and Biometricians*, Part II, Table XXIII. The majority of these values were obtained empirically from random sampling experiments. These values were of course subject to sampling error and for this reason are in three cases very slightly above d_n(max.).

Some of the preceding work was done as part of the Research and Development programme of the Ministry of Supply (S.R. 17) and appears by permission of the Chief Scientific Officer. It was completed as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

Pearson, E. S. (1942). The probability integral of the range in samples of n observations from a normal population. *Biometrika*, **32**, 301–10.

Pearson, E. S. & Adyanthaya, N. K. (1928). The distribution of frequency constants in small samples from symmetrical populations. *Biometrika*, **20 A**, 356–60.

SIGNIFICANCE TESTS FOR 2×2 TABLES

By G. A. BARNARD, *Imperial College*

Part I

The theory of statistical significance tests deals with abstractions of experimental results. The fact that the figures dealt with may happen to be tensile strengths of iron bars, or perhaps weights of babies, is ignored in the carrying out of the test; and for the purpose of statistical theory the experiment in question could just as well be represented by an experiment involving the drawing of balls from urns. In fact, it is an advantage, from some points of view, to replace the concrete experiment involved in a particular practical case by an 'abstract' urn-experiment, in order to retain in view only those features of the case which can be dealt with by statistical methods.

It is obvious enough that the first step in the statistical treatment of an experimental result may be represented as the replacement of the concrete experiment by an 'urn-experiment'; but the implications of this have not always had the continuous attention they deserve. Once the abstract picture has been formed, the analysis of it is largely a matter of pure mathematics. What distinguishes the statistician from the pure mathematician, in this connexion, should be the statistician's ability to form *valid* abstract pictures of concrete cases, and his clear recognition of the limits of validity of his abstract pictures. Yet we find relatively little discussion in statistical text-books of the process of formation of these abstract pictures.

It is the purpose of the first part of this paper to draw attention to the confusion which may arise through the possible formation of several different abstract pictures, each of which may apply to some concrete cases, though not to others.

Suppose we are given two mass-production processes, A and B, and we wish to test whether process A and process B are equally satisfactory, in the sense that neither process is more likely to produce defective items than the other. For this purpose we take, say, m articles made by process A, and n made by process B, and test them, under suitable conditions. We find that a out of the m articles are defective, while b out of the n articles are defective, a result which can be represented in the form of a 2×2 table (Table 1).

Table 1

	I (defective)	II (non-defective)	Total
Process A	a	c	m
Process B	b	d	n
Total	r	s	N

The statistical analysis of results of this type has been much discussed, but it seems to have escaped notice that, on the facts incompletely stated as above, it is possible to form several different abstract pictures, any one of which might be appropriate to the real case in question. The adoption of one picture rather than another will depend, in a given case, on further knowledge which is not specified above.

The basis of Fisher's 'exact' test

The current generally accepted test for results of the above type is that given by Fisher (1941), or some approximation to it. The simplest abstract picture* to which this test corresponds would seem to be one in which the m articles made by process A and the n articles made by process B are represented by N similar balls, m of them marked A and n marked B. The N balls are put into an urn, and then withdrawn in random order. As they are withdrawn, the balls are placed, in order, in a row of N receptacles, r of which have been marked 'I', the remainder being marked 'II'. The result of Table 1 then represents the observation that a of the balls marked A are in receptacles marked 'I'. The probability of such a result, in such an experiment is

$$\frac{m!\,n!\,r!\,s!}{N!\,a!\,b!\,c!\,d!} \tag{1}$$

which can be seen by considering that the contents of the r receptacles marked 'I' form a sample of r from an urn containing m balls marked A and n balls marked B, the sampling being done without replacement. The probability (1), added to those of all results less probable than that obtained, is the basis of Fisher's test.

In the concrete case given, the N balls, initially similar, may be taken to correspond with the N items of raw materials. The process of labelling the balls A and B corresponds to the selection of m of the items of raw material, and their fabrication into articles by process A, and the fabrication of the n remaining ones by process B. The N receptacles into which the balls are eventually placed then represent the N 'test occasions' which must be provided for when the experiment is laid out. The fact that these receptacles are labelled 'I' or 'II' before the balls are placed in them corresponds to the assumption of the hypothesis being tested—that the processes do not differ in respect of liability to defectives, so that whether or not a given article is defective has nothing to do with whether it is A or B. The labelling 'I' or 'II' is thus assumed independent of the labelling of the balls. Finally, the random allocation of balls to receptacles corresponds to a precaution which might have been taken in the concrete case, viz. the random order of test of the article secured by the use of random numbers or the like.

The basis of the C.S.M. test

Another abstract picture, also applicable to the concrete case as incompletely described above, forms the basis of the test to be developed in the later part of this paper, which we have called the C.S.M. test. In this picture, the two processes A and B, are represented by two urns, A and B, each urn containing a large number of balls, some of which are marked 'I', while the others are marked 'II'. The selection for test of m articles of process A is represented by the random drawing of m balls from urn A; and similarly for the n articles of process B. The test procedure corresponds to the examination of the balls, to see whether they are marked 'I' or 'II'. The liability of process A to produce defectives is represented by the proportion p_a of balls marked 'I' in urn A, while p_b similarly represents the liability of process B. The hypothesis we wish to test says that $p_a = p_b = p$, say. The probability of a result such as that of Table 1 is very nearly

$$\frac{m!}{a!\,c!}p_a^a(1-p_a)^c \times \frac{n!}{b!\,d!}p_b^b(1-p_b)^d \tag{2}$$

* Though not the only possible one. By following Fisher's argument, as given in his book, one can construct a more complicated picture which leads to a similar result.

which, on the hypothesis tested, becomes

$$\frac{m!\,n!}{a!\,b!\,c!\,d!}p^r(1-p)^s. \tag{3}$$

We may notice that the expression (3) differs from (1) by a factor

$$\frac{N!}{r!\,s!}p^r(1-p)^s$$

and it would have been obtained in the earlier case if we had assumed that the labelling of the receptacles was itself done randomly, by selection of N labels from a box containing a large number of labels, the proportion marked 'I' being p.

To justify the application of our second picture to a concrete case, we should have to be satisfied that the conditions of process A and those of process B were sufficiently stable, in a statistical sense, to justify the formation of the notions corresponding to p_a and p_b. We should further have to make sure that our selection of samples of m and n respectively was for practical purposes random. And finally, we should have to be reasonably sure that the conditions of test themselves had practically no influence on the results of the test—that the test used revealed a real property of the article tested, rather than a property of the individual conditions of test.

Another type of abstract experiment

Another case of common occurrence may be represented by a single urn, containing balls each of which carries two marks—one mark being either A or B, the other mark being either 'I' or 'II'. The experiment consists in drawing N balls from the urn, at random, and examining their markings. If the proportion of balls marked 'A I' is p_{a1}, while p_{b1}, p_{a2}, p_{b2} similarly represent the proportions of the other markings in the urn, the probability associated with Table 1 in this case is

$$\frac{N!}{a!\,b!\,c!\,d!}p_{a1}^a p_{b1}^b p_{a2}^c p_{b2}^d \tag{4}$$

by the multinomial theorem, provided the number of balls in the turn is large. In this case the hypothesis tested, that the markings 'I' and 'II' on the one hand, and the markings A and B on the other, are independent, may be put in the form

$$p_{a1}p_{b2} = p_{a2}p_{b1}$$

and, assuming that $(p_{a1}+p_{a2}) = p'$ and $(p_{b1}+p_{b2}) = 1-p'$, and $(p_{a1}+p_{b1}) = p$ and $(p_{a2}+p_{b2}) = 1-p$, do not vanish, the probability of our result, on the hypothesis tested, can be expressed as

$$\frac{N!}{a!\,b!\,c!\,d!}p^r(1-p)^s(p')^m(1-p')^n \tag{5}$$

which differs from (3) by a factor

$$\frac{N!}{m!\,n!}(p')^m(1-p')^n.$$

This shows that (5) is related to (3) in much the same way as (3) is related to (1).

This situation could present itself in our concrete case if the articles made by the two processes A and B were mixed up together in a common store, and the test sample of N were randomly drawn from this store, the subsequent conditions being as in the second case. Statisticians with industrial experience may perhaps feel it is unlikely that the experiment

would be performed in this way; but it must be admitted that it could have been. Cases such as this seem to occur more frequently in biometric investigations, where a population of animals is being tested for the association or otherwise of two characters.

Nomenclature

The name 'double dichotomy' has been applied generally to all experiments leading to results of the form of Table 1, but the foregoing analysis would suggest that it might be more appropriate to restrict this term to the third case we have indicated. Since the second case can be obtained from the third by supposing the numbers of articles made by process A and by process B to be fixed, we might then call the second case the (singly) restricted double dichotomy. Similarly, the first case would be called the doubly restricted double dichotomy. Such a nomenclature, apart from a lack of euphony, would be open to the objection that it would tend to imply that the third case was the general one, the first two being derivatives of it. This, in turn, would imply that the subject-matter of our investigation in cases one and two was in reality a four-fold universe, the restrictions on numbers being merely matters of experimental technique. But such is not always the case. The question implied in our second case presupposes two two-fold populations, which are to be compared, and no four-fold super-population need exist for this question to have meaning.

We therefore propose the names 'double dichotomy' for the third case, '2×2 comparative trial' for the second case, and '2×2 independence trial' for the first case, though here again an objection on aesthetic grounds would be easy to sustain.

Finer distinctions

In principle it could be maintained that there is a distinction between the 2×2 comparative trial, as instanced above, and a restricted double dichotomy. As we have said, the fundamental subject-matter of a 2×2 comparative trial is a pair of populations; while the subject-matter of a restricted double dichotomy is a four-fold population from which we happen, by an accident of experimental technique, to be able to extract samples in which the numbers of items having certain characteristics are fixed. The latter case could arise, for example, if an attempt was being made to discover association between colour of eyes in school-children and some less easily identified characteristic, such as membership of a particular blood-group. We could imagine that an experimenter might pick out m children with (say) blue eyes, and n without blue eyes, and then, having obtained his samples, he might subject them to a test for blood-group. The conclusions drawn from such an experiment would presumably be intended to apply to the population of school-children, a four-fold one relative to the two characteristics in question. The distinction between the two cases comes out if we consider what happened if, in the 2×2 comparative trial, all items tested turn out to be defective. In this case we should say that our question, whether $p_a = p_b$ or not, tends to be answered in the affirmative. In the case of the school-children, if they all turn out to have the same blood-group, then no conclusion on our question about the four-fold population can be drawn at all.

Similar distinctions apply to the 2×2 independence trial. In the psycho-physical experiment described by Fisher (1942), where the point at issue is whether or not a lady can tell whether the milk or the tea has been put in the cup first, no statistical population is presupposed. The question would have meaning even if we refused to regard the order of insertion of milk or tea as ever being a matter of chance, while at the same time we regarded

the lady's guess as equally determinate. The 'statistical population' enters into this experiment only in the experimental technique, via the randomization procedure used to fix the order of presentation of cups; it does not enter into the question being asked. In this case, the extreme result, in which in fact the milk was put in first every time, while the lady guessed every time that it was otherwise, would be taken as evidence against the lady's claim. But such a result could by itself have no meaning for the question asked in the case of a restricted 2×2 trial or a doubly restricted double dichotomy.

Further types of experimental procedure leading to results expressible in the form of Table 1 are the various sequential procedures that have been described for deciding questions of the kind we have been discussing (3, 4). Yet another procedure is one where the conditions of trial vary from one block of tests to another—as when an open-air trial runs over several days of inconstant weather. Here we might suppose there were k pairs of urns, (A_1, B_1), (A_2, B_2), ..., (A_k, B_k). The distinctions here are, however, obvious enough, and they are worth noting only in order to emphasize that the mere fact there results are presented in the form of Table 1 is not in itself sufficient to specify an appropriate test of significance.

PART II

The significance test for the 2×2 *trial*

Roughly speaking, the object of a significance test as applied to results of the type considered, is to answer the question: Can these results be ascribed to 'chance'? In this form, the question is not sufficiently precise. If our 'urn model' for the 2×2 comparative trial is adequate to represent the experiment actually carried out, then the results will in any case be 'due to chance', in some sense. What we wish to know in this case is whether a particular kind of chance—namely, one in which $p_a = p_b = p$—can be said to account for our results. If the results are such that this explanation of them is untenable, then we may conclude either, that our particular 'urn model' of the experiment is inadequate anyway; or we may retain the model, and conclude that p_a and p_b must be unequal. In most cases, of course, we shall reach the latter conclusion, since we would not have made up the urn model in question unless we had some reasons for believing in its adequacy; but it is well to bear in mind the first alternative, in case a re-examination of the circumstances may make us change our minds. A point very strongly emphasized by Fisher in his book *The Design of Experiments* is, that we ought to have in mind a particular 'urn model' *before* the experiment is performed, and arrange the conduct of the experiment so that the adequacy of this urn model is not likely to be questioned afterwards.

With the qualifications indicated, we can say that the object of the significance test we propose to develop is, to enable a particular class of explanations of our experimental results to be ruled out as untenable. Specifically, given results like those of Table 1, we want to be able to say that they could not be accounted for by supposing that the experiment we actually performed was analogous to the urn experiment with two urns in which $p_a = p_b = p$. This raises the question, in what sense could such a supposition fail to account for the observed results? Any result of the form of Table 1 *could* arise in an experiment of this kind, when our supposition is true. Why, then, should we select some results of this form and say they are incompatible with our supposition?

In the last analysis, this question cannot be answered without an examination of what is meant in general bv statements involving probabilities, a point which is still the subject of

controversy. But in our particular case (if not in all cases) we can avoid giving a general answer to the question of what probability is, by considering the practical circumstances which form the setting for our particular problem, and the uses to which we propose to put the answer. In fact, in our case we are interested in the equality or otherwise of p_a and p_b because we want to decide which of the two processes, A and B, is to be preferred, from the point of view of defectives produced. To say that p_a is greater than p_b will mean, for us, that process B is preferable, and conversely if p_b is greater than p_a, while to say that p_a and p_b are equal will mean that there is nothing to choose between the two processes. In fact, to say that $p_a = p_b$, in our case, means that, if process A and process B are both used, then it will be found that the frequencies with which defectives appear in the two processes will, for practical purposes, be equal.* Thus we shall assert that results in which the observed frequencies, a/m and b/n, differ widely, are incompatible with the supposition that $p_a = p_b$; in doing so, we shall be neglecting as impossible a class of events which are in reality logically possible, but whose probability is small. The precise formulation of a test of significance then reduces to a precise formulation of what is meant by a 'wide difference' in the frequencies a/m and b/n, and to an evaluation of the probability of those events which are being neglected as impossible.

The lattice diagram

If we consider the first problem, of arranging results like those of Table 1 in order of the relative 'width' of the differences they indicate, a first step is the enumeration of all possible results in a convenient form.

Logically, we should begin by noting that Table 1 is really an abbreviated version of the results of any one particular experiment, which will to start with be like those of Table 2 (where we have taken $m = 8$, $n = 6$, for definiteness).

Table 2

Urn:	A	A	A	A	A	A	A	A	B	B	B	B	B	B
Mark:	II	I	II	II	I	II	II	II	I	I	II	I	I	I

But if, as we are presupposing, our urn analogy is adequate to represent the conditions of the experiment, the order in which the results were obtained must be irrelevant to the interpretation of results. If the conditions of trial varied during the course of the experiment, this assumption might not be correct—for example, if the trial were an open-air trial, and it began to rain half-way through. We are assuming that the urn analogy is adequate, and so we must treat all results like Table 2 which give the same values to a, b, c, d, in Table 1, as equivalent. Table 1 therefore stands for $m!\,n!/a!\,b!\,c!\,d!$ distinct, but equivalent, results which we shall not distinguish from now on.

If we now take rectangular axes in a plane, we can represent Table 1 by the point whose coordinates are (a, b). Thus 'x' in Fig. 1 represents the set of results equivalent, in the sense of the previous paragraph, to the results of Table 2. At the same time, all possible results of the experiment which gave rise to Table 2 are represented by the points of the rectangle

* We hope that the qualifications we have attached to our statements will be sufficient to guard us against the accusation that we have adopted in full a 'frequency theory' of probability. The frequency interpretation is relevant to our particular problem; other problems may involve other interpretations. More than one interpretation may be relevant in a single problem.

$PQRS$. We call this representation of possible results the lattice diagram.* Our problem may now be regarded as one of ordering the points of the lattice diagram according to the 'width' of the difference they indicate.

Conditions S and C

In trying to make the idea of 'width' of difference precise, we are up against difficulties similar to those attaching to the interpretation of results on the basis of incomplete information about the circumstances of the experiment. The information given at first was compatible with several distinct 'urn models'. Similarly, the information given now is compatible with several different notions about 'width' of difference. We may be concerned with the arithmetical size of the difference $p_a - p_b$, or with the ratio p_a/p_b, or with the logarithm of this ratio, or with some more complicated function.

Logically, therefore, we should expect to set up various tests, based on various ideas of what constitutes 'width' of difference in probability or frequency (in Neyman and Pearson's

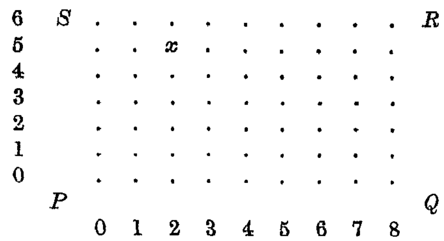


Fig. 1

Table 3

	I	II	Total
A	c	a	m
B	d	b	n
Total	s	r	N

language, corresponding to various weight functions over the space of alternatives to the hypothesis tested). But here a factor which may simply be described as laziness enters in. If we carried our ideas to their logical conclusion, we should find ourselves constructing a new test for almost every new experiment we had to deal with; and the time and effort involved in this are too great. Consequently, we confine our attempt to producing a test which will be reasonably applicable to a wide class of cases of the type specified, without suggesting that this test is unique, or 'best possible'.

First, then, in our ordering of points in the lattice diagram, we propose that the same rank should be given to the point $((m-a), (n-b))$ as to the point (a, b). This condition we propose to call the 'symmetry condition', or 'condition S'. It amounts to saying, that if Table 1 is to be considered as indicating a real difference between p_a and p_b, then so is Table 3, in which the labels 'I' and 'II' have been interchanged. If, when we are testing whether

* Not the sample space of Neyman and Pearson. In the sample space, different results equivalent to Table 2 are represented by different points.

$p_a = p_b$, we can say we are also testing whether $1-p_a = 1-p_b$, from the same point of view, then this symmetry condition is clearly justified.*

Next, we propose that in our ordering, the two points which, respectively, have the same abscissa or the same ordinate as (a, b), and which lie further from the diagonal PR, shall be considered as indicating wider differences than (a,b) itself. Thus, referring to Fig. 1, the points immediately above and immediately to the left of the point 'x' are reckoned to indicate wider differences than the point 'x' itself. This condition implies that the set of points indicating differences as wide or wider than (a, b) will have a shape property vaguely related to convexity, and we call it the 'C condition'. It means that if we consider the table corresponding to Table 2, with cell frequencies

$$\begin{matrix} 2 & 6 \\ 5 & 1 \end{matrix}$$

as significant evidence of difference, then we must also consider the tables

$$\begin{matrix} 1 & 7 \\ 5 & 1 \end{matrix} \quad \text{and} \quad \begin{matrix} 2 & 6 \\ 6 & 0 \end{matrix}$$

as significant evidence of difference. It is difficult to imagine circumstances where this would not be so.

Geometrically, condition S implies that we can in future restrict our considerations to points in the lattice diagram lying on or above the diagonal PR, i.e. in the triangle PRS. And condition C implies that, in this triangle, our 'width of difference' must increase as we go upwards or to the left. If horizontal and vertical axes are taken at any point X in this triangle, points in the second quadrant are associated with a wider difference than X is, points in the fourth quadrant are associated with narrower differences than X is. The relative width of differences associated with points in the first and third quadrants (excluding the axes) are now determined by the conditions C and S. The ordering generated by these conditions is thus a partial, not a total, ordering; it is, in fact, a kind of conical order, in the sense of A. A. Robb. We must introduce some further condition to make the ordering total.

Probability considerations

In many simpler cases, it is possible to distinguish those events which are considered incompatible with a given probability hypothesis by their relatively low probability, compared with other possible events. Such a simple comparison of probabilities is not open to us in this case, because to each point (a, b) we have, on the hypothesis tested, associated a function

$$W(a, b; p) = \frac{m!\,n!}{a!\,b!\,c!\,d!} p^r (1-p)^s$$

which contains the 'nuisance parameter' p If we consider the relative position, in our ordering, of another point, (a', b'), we have to consider the inequality

$$W(a, b; p) < W(a', b'; p), \tag{6}$$

the truth or falsehood of which depends, in general, on the unknown p; and there is nothing in the statement of the problem, nor in the experimental method, to justify any particular choice for the value of p.

* Cases where $p_a > p_b$ is impossible are hereby neglected, strictly.

If $(a+b) = (a'+b')$, the validity or otherwise of the inequality (6) is independent of p. Thus, using this inequality as a criterion for ordering our points, we can say that in the triangle PRS, the 'width of difference' must increase as we move north-west. But this is all that can be derived from this criterion, and it is clearly even less helpful in ordering the points than the conditions C and S are. Moreover, if we recall that each point (a, b) in the lattice diagram really represents a set of $m!\,n!/a!\,b!\,c!\,d!$ distinct results, each with probability $p^r(1-p)^s$, the criterion (6) loses its plausibility.

We might try to improve the situation by associating the function $W(a, b; p)$ with a number, depending on a and b only. For fixed a and b, this number would be a functional of $W(a, b; p)$. We should clearly require that, if the inequality (6) is true for all p, then the corresponding inequality should be true of the numbers associated with $W(a, b; p)$ and $W(a', b'; p)$. The simplest functionals which satisfy this condition will be the mean value,

$$w(a, b) = \int_0^1 W(a, b; p)\,dp,$$

the maximum value

$$w'(a, b) = \max_{0<p<1} W(a, b; p),$$

and one single value

$$w''(a, b) = W(a, b; p_0).$$

Circumstances could be imagined in which any of these three criteria might produce reasonable tests of significance. For example, in certain genetical experiments we may have reason to suppose that the value $p = 1/3$ would occur more often than any other value. In such a case we might use w'', with $p_0 = 1/3$. But for general purposes taking $p_0 = 1/3$ could not be justified.

We might again argue that taking w as our criterion would correspond with the assumption that all values of p were a priori equally likely. But some would say that such an assumption was never justified; while those who would admit the assumption would in strictness do so only if we really did know *nothing* about the value of p. And in the general circumstances we are trying to cater for, we may sometimes know something vague about the value of p—such as, for example, that p will be less than $\frac{1}{2}$.

Neyman and Pearson have shown that the likelihood ratio, which in our case comes to be

$$\frac{m^m n^n r^r s^s}{a^a b^b c^c d^d N^N}$$

very often gives a good basis for ordering experimental results. We feel, however, that the criterion we shall describe in the next section has a slightly more direct justification than the likelihood ratio, though the choice, is, admittedly, largely a matter of taste.

The maximum condition

Before setting out the final condition which, with conditions S and C, will be used eventually to arrange the points of the lattice diagram in order of 'relative width of difference indicated', we need to consider the assignment of significance levels to various results.

When we say that a given result is not significant on, say, the 5 % level, we mean that such a result, or one indicating a wider difference, could occur, with probability at least 0·05, even when $p_a = p_b$. We could believe in a theory that $p_a = p_b$, without having to suppose that an event belonging to a class whose joint probability was less than 0·05 had occurred. Conversely, if a result is judged significant on the 5 % level, it means that no theory which

assumed that $p_a = p_b$ could account for the result obtained without supposing that an event of a type whose probability was less than 0·05 had occurred on the occasion in question.

Let us now consider a specific case, in which we choose numbers which in practice would be ridiculously small in order to save arithmetic. Suppose, in fact, $m = n = 2$, while $a = 2$ and $b = 0$. It follows from conditions S and C alone that in judging the significance of such a result we need consider only the probability of this result, together with its converse, in which $a = 0$, $b = 2$. If $p_a = p_b = p$, the probability of results of this type is

$$P = 2p^2(1-p)^2.$$

Now suppose that we are prepared to discard as untenable theories which require us to suppose that events of probability less than 0·05 had occurred. In such a case, we should discard a theory which supposed $p_a = p_b = 0{\cdot}1$, since in this case $P = 2(0{\cdot}1)^2(0{\cdot}9)^2 = 0{\cdot}0162$, less than 0·05. But we could not discard a theory which supposed $p_a = p_b = 0{\cdot}5$, since in this case $P = 0{\cdot}125$. In fact, our result would enable us to discard all theories involving $p_a = p_b = p$, except those for which p lay in the interval $0{\cdot}197 < p < 0{\cdot}803$. In particular practical cases we might be prepared, on grounds external to the experiment in question, to dismiss the possibility that p should lie in this interval; and in such cases we should be entitled to say that the result excludes the possibility that $p_a = p_b$.

It is easy to see that the above specific case is typical. Any set of points in the lattice diagram, considered by some criterion agreeing with conditions S and C to indicate differences as wide or wider than those of a given result, will be associated with a probability P, on the assumption $p_a = p_b = p$; and this P will be a function of p, rising from zero when $p = 0$ to a maximum in the neighbourhood of $p = \frac{1}{2}$, and then falling again symmetrically (by the S condition) to zero again at $p = 1$, somewhat as in Fig. 2. The given result by itself

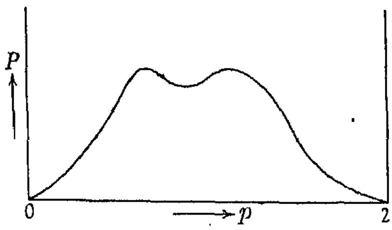


Fig. 2

will exclude the possibility $p_a = p_b$ altogether, only if the significance level adopted is greater than P_m, the maximum value of P. If our significance level corresponds to a probability less than P_m, then all we can say is, that our result is incompatible with $p_a = p_b$ unless their common value lies in a certain subset of the range (0, 1). We may or may not exclude these latter possibilities on other grounds.

In trying to construct our test, however, we have set ourselves the task of evaluating the evidence *provided by our experiment alone* in relation to the hypothesis $p_a = p_b$. It now appears that this is impossible so long as we restrict ourselves to the form, usual in such cases, of a simple statement that a given result is, or is not, significant on a given level. We have two alternatives. Either we can find an entirely new form of statement to convey what we wish to express; or we can adhere to the form of statement, and try to make the situation fit the form as nearly as possible. Perhaps the day will come when experimenters do not require answers in the form of numbers, when they are sufficiently versed in generalized mathematical analysis to be content with a function (such as the function $P(p)$), instead of a single

number. But we have not yet reached this stage; and so we propose to take up the latter alternative, and try to make the situation fit the standard form of statement of significance tests as nearly as possible.*

Our difficulty arises from the dependence of P on p. If the graph of P against p were a horizontal straight line, our difficulty would be overcome. What we propose, therefore, is to try to make the graph of P against p as near to a horizontal line as possible, by suitably adapting our idea of what is meant by 'width of difference'. In making this adaptation, we shall secure that we do not violate the common-sense requirements as to the meaning of the term 'width of difference', by requiring that conditions C and S should always be satisfied.

The maximum condition

The condition C requires that, of all points in the triangle PRS, that indicating the 'widest difference' must be the point S at the corner (Fig. 1). The function P associated with this point and its converse, Q, which we may denote as $P(0,6;p)$, is

$$P(0,6;p) = p^6(1-p)^8 + p^8(1-p)^6$$

and the maximum P_m occurs here when $p = \frac{1}{2}$, where we have

$$P_m(0,6) = 1/2^{13} = 1{\cdot}22 \times 10^{-4}.$$

The condition C requires that the only points which might be considered as coming next after S, in order of decreasing 'width of difference' are $(1,6)$ and $(0,5)$. We have to adopt some principle to choose between these two.

If $(1,6)$ were taken next after $(0,6)$, the function P associated with it would be

$$P'(1,6;p) = P(0,6;p) + 16p^7(1-p)^7$$

and $P'_m(1,6)$ would come to $9/2^{13} = 10{\cdot}97 \times 10^{-4}$. On the other hand, if $(0,5)$ were chosen next, instead of $(1,6)$, we should have

$$P(0,5;p) = P(0,6;p) + 6[p^9(1-p)^5 + p^5(1-p)^9]$$

and $P_m(0,5)$ would come to $8{\cdot}58 \times 10^{-4}$, the maximum occurring when $p = \frac{1}{2} \pm \frac{1}{2}\sqrt{(6/70)}$. Thus $P_m(0,5)$ is smaller than $P_m(1,6)$, and this lower maximum is associated with a flatter curve of $P(0,5;p)$. Since a flat curve is our aim (the horizontal line being the ideal), we choose $(0,5)$ as the point to come next after $(0,6)$, rather than $(1,6)$.

Having chosen $(0,5)$ as the next 'widest difference' point, the C condition restricts us to the points $(1,6)$, and $(0,4)$, as candidates for the next position. We consequently compare

$$P(1,6;p) = P(0,5;p) + 16p^7(1-p)^7$$

with
$$P''(0,4;p) = P(0,5;p) + 15[p^4(1-p)^{10} + p^{10}(1-p)^4]$$

and the lower value of P_m as criterion shows that $(1,6)$ is now to be taken. At the next stage, we shall have to compare the functions associated with $(0,4)$, $(1,5)$ and $(2,6)$. In this way we can arrange the points of the lattice diagram in order, step by step.

The principle involved, which we call the 'maximum condition', may be formally stated as follows:

Considering only points for which a/m is less than b/n, if the first $(n-1)$ points (a_1,b_1), (a_2,b_2), ..., (a_{n-1},b_{n-1}), in order of decreasing 'width of difference' have been chosen, and

* In the example just taken we might make a kind of 'conditional confidence interval statement', that, if p existed, we should have $0{\cdot}197 < p < 0{\cdot}803$ with confidence coefficient $0{\cdot}95$.

(a_{n-1}, b_{n-1}) is associated with the function $P(a_{n-1}, b_{n-1}; p)$, then the nth point, (a_n, b_n) is that point, of all points (a, b) permitted by the C condition, for which

$$P_m(a, b) = \max_{0<p<1}\left[P(a_{n-1}, b_{n-1}; p) + \frac{m!\,n!}{a!\,b!\,c!\,d!}(p^r(1-p)^s + p^s(1-p)^r)\right]$$

is least. (a_n, b_n) is then associated with the function

$$P(a_n, b_n; p) = P(a_{n-1}, b_{n-1}; p) + \frac{m!\,n!}{a!\,b!\,c!\,d!}[p^r(1-p)^s + p^s(1-p)^r].$$

To complete the specification of the ordering, we have to legislate for the case where there are several points giving the same value of $P_m(a, b)$, this value being less than that associated with any other permissible point. In this case we lay down that all such points are to be given the same rank, and the second term in the expression for $P(a_n, b_n; p)$ is to be replaced by the corresponding sum over all these points. If there are k such points at any stage, then the next point after them will be denoted as the $(n+k)$th point in the ordering. This requires, for example, when $m = n$, that the points (a, b) and (b, a) are always to be taken together.

Finally, the significance level to be attached to the point (a_n, b_n) will be

$$P_m(a_n, b_n) = \max_{0<p<1} P(a_n, b_n; p).$$

This guarantees that our test will be a 'valid' one, in the sense that, if we judge a result incompatible with the hypothesis $p_a = p_b$; on a given level of significance, then *all* the possibilities of the form $p_a = p_b$ are excluded, to the given level. Thus no further information, external to the experiment in question, could make us decide that a result judged significant by our test was not in fact so (holding, of course, to a fixed significance level); on the other hand, we still have the possibility that other information may lead us to consider as significant results which appear in themselves not to be so. The formulation of our maximum condition is made so as to minimize this latter possibility. Our test is thus conservative, in the sense that we do not draw the conclusion $p_a \neq p_b$ unless this is certainly warranted by the data; but it might be called 'progressive conservative', because, of all such conservative tests, it will be the least conservative.

Another aspect of the maximum condition

When the author first approached the problem of analysis of experimental results of the type now considered, he did so from the point of view of regarding the significance level to be used as being fixed in advance, say at the 5 % level. From this point of view, the problem of constructing a test resolved itself, not into one of *ordering* the points in the lattice diagram, but into one of choosing a region, or set of points in the lattice diagram, such that any point belonging to this region could be regarded as evidence of inequality of p_a and p_b, on the given level of significance. The condition of symmetry required that such a region should consist of two similar parts, one above the diagonal PR, and one below it. The condition C required that the part of the region lying above the diagonal PR should be so shaped that if a point X belonged to the region, then so would all points lying north or west of X. There remained the problem, to decide which of the many regions satisfying these two conditions should be the one adopted.

To settle this, to any such region R we can associate a function

$$P(R; p) = \sum_{(a,b)\in R} \frac{m!\,n!}{a!\,b!\,c!\,d!} p^r(1-p)^s$$

and such a region will give a 'valid' test of significance provided that

$$\operatorname*{Max}_{0<p<1} P(R; p) \leqslant 0{\cdot}05.$$

There will not be so many regions satisfying this validity condition as well as the conditions S and C. We proposed, therefore, to select that region from among these, which had the greatest number of points in it. This last condition was what we then called the 'maximum condition'. The fact that this region would not be unique in cases where $m = n$ was taken care of by requiring a subsidiary symmetry condition that in such cases (a, b) and (b, a) should always be taken together.

What we have now adopted as the 'maximum condition' can be seen to be related to this earlier version, by the consideration that, roughly speaking, apart from effects due to the discreteness of the lattice diagram, holding the number of points in the region constant, and then choosing the region which gives the lowest value for P_m, as we do now, comes to the same thing as holding P_m constant, and then choosing the region to have the maximum number of points.

Other things being equal, the 'power' of a test, in the sense of Neyman and Pearson, will increase with the 'volume' of the rejection region chosen. In this sense we can say, roughly, that the maximum condition secures that our test should be as powerful as possible, consistent with validity.

Practical formulation of the test

Some statistical tests (such as that due to Fisher, already mentioned), can be carried out in the form of a direct calculation from the data, without reference to any special tables. Most other tests require the use of special tables which, however, are for the most part tables of single or double entry, perhaps triple entry, if the level of significance is regarded as a variable. In our case, regarding the level of significance as a variable, a table of quadruple entry would be required.

Ideally, a set of tables, one for each pair of values of m and n $(m \geqslant n)$ would be required. The table would be in the shape of a right-angled triangle, corresponding to the triangle PRS of Fig. 1, and divided into squares, each square corresponding to given values of a and b. Within each square (a, b) would then appear a number, the value of $P_m(a, b)$. This value of $P_m(a, b)$ then is the maximum probability of obtaining the result (a, b), or one indicating a wider difference, if $p_a = p_b$. A comparison of $P_m(a, b)$ with the significance level adopted will then decide the significance or otherwise of our result. In any particular case we shall be able to see which tables, in the sense of our test, are regarded as indicating a wider difference, by noting which points are associated with lower values of $P_m(a, b)$.

In practice, it will be impossible to construct such tables for a large range of values of m and n. But for larger values of m and n, a test based on a normal approximation to the distributions involved will be quite adequate for practical purposes. In fact, the test we have proposed will itself approximate, in some sense, to a test based on the normal distribution, though we do not enter into a detailed discussion of the relationship between the two tests here.* Tables are thus required for our test only for small values of m and n. In spite of advice by statisticians to the contrary, such small values of m and n continue to occur frequently in practice.

* The general question of the sense in which tests are regarded as 'asymptotically approaching' normal tests is a subject for another paper. Professor Pearson's paper which follows, bears on this point.

In the Appendix we give specimen tables for the cases where $N = 14$. The comparative figures for the Fisher test, also given in the Appendix, indicate that the differences between the two tests are appreciable. An exploration is now under way into larger values of m and n, and it is hoped to report on this in due course.

Other applications of the C.S.M. procedure

We have spoken of our test as *the* C.S.M. test, as if the case dealt with above were the only case to which the procedure adopted was applicable. But similar methods could be used in many other cases. In particular, a method closely following the one we have used might be applied to the case we have called the double dichotomy, which differs from the 2×2 comparative trial in that two 'nuisance parameters', p and p' are present, instead of only one. The 2-dimensional lattice diagram of the 2×2 trial is replaced by a 3-dimensional regular tetrahedron of points with homogeneous coordinates (a, b, c, d), connected by the relation

$$a+b+c+d = N.$$

Two opposite edges of this tetrahedron correspond to $m = 0$ and $n = 0$, and sections of the tetrahedron by planes parallel to these edges will look exactly like lattice diagrams for the 2×2 case and within these sections, relative probabilities will behave just as in the 2×2 case. An examination of the possibilities, however, indicates that not much is to be gained by a detailed treatment. The C.S.M. test for 2×2 comparative trials will be a valid test if applied to double dichotomies. It will err somewhat on the side of 'conservatism', but the error does not appear to be large, except when the numbers involved are exceedingly small.

It is with a view to further applications of the approach used in this paper that we have retained the C condition as a separate requirement, although it is easy to see that it could be absorbed into the M condition as we have given it.

In writing this paper the author has had great personal help and encouragement from Prof. E. S. Pearson, to whom he wishes to express his very deep thanks.

Summary

In Part I we discuss various types of experiment, each of which may give rise to results in the form of a 2×2 table. It appears that significance tests which may be appropriate for one type of experiment will not necessarily be appropriate for another.

In Part II a test is developed for experiments of the type called '2×2 comparative trials'.

APPENDIX

Tables for the CSM test

Three tables are given below to illustrate the application of the ideas given in the main paper to the construction of a test for 2×2 comparative trials. The cases covered are pairs of samples, sizes (7, 7), (8, 6), and (9, 5). The small figures in brackets in the (7, 7) table gives significance levels on Fisher's 'exact' test for 2×2 independence trials, for comparison. Only half of the (8, 6) and (9, 5) tables are given; the missing parts can be filled in by symmetry. The following examples show the meaning and use of the tables:

Example 1. Two boxes, each containing a large number of components, are to be tested for comparative quality measured by the respective proportions of defective components they contain. Two samples, each of seven components, are taken, at random, one from each box. One sample gives four defectives, the other, none. What is the significance of this result, in relation to the hypothesis that the boxes have the same quality?

Answer. Entering the (7, 7) table at the point (0, 4), we find the number 2·4. This means that the result is evidence against the hypothesis, on the 2·4 % level of significance.

Table for $m = n = 7$

7	0·012 (0·058)	0·18 (0·23)	0·70 (2·1)	2·4 (7·0)	7·5 (19)	20 (46)	—	—
6	0·18 (0·23)	1·3 (2·9)	5·7 (10)	13 (27)	—	—	—	—
5	0·70 (2·1)	5·7 (10)	21 (29)	—	—	—	—	20 (46)
4	2·4 (7·0)	13 (27)	—	—	—	—	—	7·5 (19)
3	7·5 (19)	—	—	—	—	—	13 (27)	2·4 (7·0)
2	20 (46)	—	—	—	—	21 (29)	5·7 (10)	0·70 (2·1)
1	—	—	—	—	13 (27)	5·7 (10)	1·3 (2·9)	0·18 (0·23)
0	—	—	20 (46)	7·5 (19)	2·4 (7·0)	0·70 (2·1)	0·18 (0·23)	0·012 (0·058)
	0	1	2	3	4	5	6	7

More precisely, what is asserted is, that the maximum probability of getting a result not less significant than that obtained, is 0·024. And the results which are not less significant are those which correspond to points in the table with numbers not greater than 2·4, viz. (0, 4), (7, 3), (0, 5), (7, 2), (0, 6), (7, 1), (0, 7), (7, 0), (1, 6), (6, 1), (1, 7), (6, 0), (2, 7), (5, 0), (3, 7), (4, 0). By suitable choice of the proportion defective, we could construct a pair of boxes, of equal quality, which would give samples falling in this group 24 times out of 1000, on the average; but we could not, by any choice of proportion defective, retain equal quality and yet have results in this group more often than 24 times in 1000.

Table for $m = 8$, $n = 6$

	0	1	2	3	4	5	6	7	8		
6	0·012	0·18	0·71	2·5	5·3	13	—	—	—	—	
5	0·085	1·3	6·6	11	—	—	—	—	—	—	5
4	0·44	3·9	19	—	—	—	—	—	—	—	4
3	1·9	.16	—	—	—	—	—	—	—	6·3	3
2	8·0	—	—	—	—	—	—	20	3·8	0·86	2
1	23	—	—	—	—	—	14	7·4	1·3	0·13	1
0	—	—	—	16	10	5·3	2·3	0·62	0·19	0·012	0
	0	1	2	3	4	5	6	7	8	9	

Table for $m = 9$, $n = 5$

Example 2. The situation is as before, except that the first sample has nine components, none of them defective, while the second sample has five components, four of them defective.

Answer. Here, to use the table as given, we have to compare numbers effective, rather than numbers defective—viz. we consider the pair (9, 1) rather than (0, 4). Entering the (9, 5) table at (9, 1) we find 0·13. The result is evidence against the hypothesis of equal quality, on the 0·13 % level of significance.

Thanks are due to Miss Lang, who has checked the computations.

REFERENCES

Fisher, R. A. (1941). *Statistical Methods for Research Workers*, 8th ed. Edinburgh: Oliver and Boyd.
Fisher, R. A. (1942). *The Design of Experiments*, ch. ii. Edinburgh: Oliver and Boyd.

THE CHOICE OF STATISTICAL TESTS ILLUSTRATED ON THE INTERPRETATION OF DATA CLASSED IN A 2×2 TABLE

By E. S. PEARSON

CONTENTS

(i) Introductory

1. The problem of testing the significance of a difference between two proportions is one which receives early attention in text-books on mathematical statistics, and it might be thought to be one of the questions whose final solution lies behind us. It is a problem whose simplicity makes it easy to examine the logical cogency of the methods put forward for its solution, but, on examination, it is evident that they have not yet been rounded off satisfactorily. The origin of the present paper lies partly in an investigation commenced in 1938 and discussed at the time in College lectures, and partly in recent correspondence in *Nature* in which G. A. Barnard (1945*a*, *b*) and R. A. Fisher (1945*a*) have taken part.* This correspondence has suggested that in a problem of such apparent simplicity, starting from different premises, it is possible to reach what may sometimes be very different numerical probability figures by which to judge significance.

2. Such a difference in levels of significance in the solution of an everyday problem is obviously puzzling to the users of statistical methods who are accustomed to accept the technique as an established procedure and have not the opportunity for a critical examination of the conditions under which probability theory is brought to bear as a guide to action. For the question here at issue is a fundamental one of why and how our judgement is influenced by the calculation of a probability, and the dilemma raised by the Barnard-Fisher correspondence can only be answered in terms of our views on the practical function of the theory. We may all agree that in practice we use probability figures derived from an analysis of numerical data to help us to make up our minds on the next step, whether in experimental research or executive action. But what form of presentation of the probability set-up is likely to result in the greater number of sound decisions is likely to be always a matter for differences of opinion.

3. All that I can do is to approach the problem of the 2×2 table from the viewpoint which appears most helpful to me. In the preceding paper Mr Barnard has elaborated the

* There was also an earlier discussion on the same subject between E. B. Wilson (1941, 1942) and R. A. Fisher (1941).

views expressed in his letters to *Nature*. Such discussion is, I believe, desirable, even though controversial issues are raised. For the value of the whole elaborate structure of the modern theory of mathematical statistics depends at least in part on the sense in which the individual statistician appreciates the meaning of the probability model he is using when drawing the practical conclusions from his analysis of data. I have used the words 'in part', for it is true that the analytical process of applying the statistical technique to experimental data may in itself be enormously illuminating even without paying any close regard to a final probability figure. Such is the case, for example, with the technique of analysis of variance, where the mere process of breaking up a total sum of squares into parts with which different sources of variability can be associated, brings with it a reward in clear thinking even without the application of a probability test.

4. *There is a very wide variety in the types of situation in which probability theory is* introduced to help in reaching a decision as to further action.

(A) At one extreme we have the case where repeated decisions must be made on results obtained from some routine procedure carried out under controlled conditions.

(B) At the other is the situation where statistical tools are applied to an isolated investigation of considerable importance in which many of the issues involved in the conclusion can hardly be assessed in numerical terms.

5. Two situations of this kind, in which the statistical technique involved is that of testing the significance of a difference between two proportions, may be illustrated from problems arising in the 'proof' of armour-piercing shot or shell.

6. *Example of type A*. In the proof of small anti-tank, armour-piercing shot it might be decided to set aside, as a standard, a batch of shot whose quality has been established by special trials; against this standard, later batches can be compared. The variable measured is the proportion of shot which fail to perforate a plate of specified thickness when fired with a given striking velocity. The use of standard shot is necessary for calibration purposes, because there are inevitable changes in toughness from one proof plate to another and only a limited number of shot can be fired at a single plate. Then the situation might be summed up as follows:*

Aim of proof. To ensure that as few batches as possible are passed into service which are less effective than the standard.

Method of proof. Twelve rounds of the standard and twelve of the batch under test to be fired, round for round, against a single test plate and a record kept of the number of failures in each group, say a and b.

Routine sentencing rule. This should lay down a ready means of determining, from a knowledge of a and b, whether to class the new batch as inferior to the standard or not.

Assumptions accepted in using rule. That the two samples of twelve shot have each been randomly selected from the much larger batches. That against the particular plate used, a proportion p_1 of the standard and p_2 of the new batch would fail to give satisfactory perforation at the specified striking velocity. That while p_1 and p_2 would be different for other plates, if $p_2 > p_1$ for one plate, it will be so for all other plates. The objective is to segregate batches of shot for which $p_2 > p_1$.

* It has been somewhat simplified for illustrative purposes, e.g. complete control of the striking velocity is not in practice possible.

7. *Example of type B.* Two types of heavy armour-piercing naval shell of the same calibre are under consideration; they may be of different design or made by different firms. Since the cost of producing and testing a single round of this kind runs into many hundreds of pounds, the investigation is a costly one, yet the issues involved are far reaching. Twelve shells of one kind and eight of the other have been fired; two of the former and five of the latter failed to perforate the plate. In what way can a statistical test contribute to the decision which must be taken on further action?

8. In dealing with Example A the guiding principle followed in seeking help from the theory of probability can be very simple. We can set as our object a rule which:

(i) will result in an increasing chance of detecting that $p_2 > p_1$, the larger the difference;

(ii) will leave only a small chance of segregating the new batch wrongly when, in fact, $p_2 \leqslant p_1$.

Diagrammatically the rule would consist in segregating the new batch when the point (a, b) falls within some such area as that shown shaded in Fig. 1. In this problem involving a routine procedure, it is the long-run frequency of different consequences of the proof sentencing which is of importance, and probability theory is introduced to provide a measure of expected frequency. This method of introducing the theory of probability into this proof problem is not necessarily the only one that could be adopted in fixing a routine procedure, but it is a simple one and, since simplicity has the merit of appealing to the user's understanding, it has great advantages.

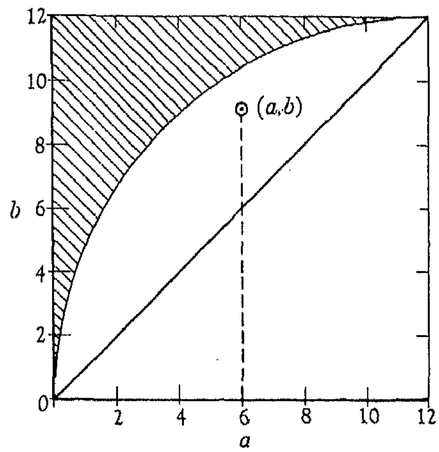


Fig. 1

9. When dealing with Example B a very considerable number of factors must be weighed in the balance, and the result of a statistical test of significance could never be the over-riding one. There will be other information as to the effect of changes in shell design, possibly from shell of different calibre; information as to the uniformity in quality of output of the firm or firms concerned; questions of cost and of general policy. He would be a bold man who would attempt to express these in numerical terms. Whereas when tackling problem A it is easy to convince the practical man of the value of a probability construct related to frequency of occurrence, in problem B the argument that 'if we were to repeatedly do so and so, such and such result would follow in the long run' is at once met by the common-sense answer that we never should carry out a precisely similar trial again.

10. Nevertheless, it is clear that the scientist with a knowledge of statistical method behind him can make his contribution to a round-table discussion, provided he has acquired a grasp of the practical issues. Starting from the basis that individual shell will never be identical in armour-piercing qualities, however good the control of production, he has to consider how much of the difference between (i) two failures out of twelve and (ii) five failures out of eight is likely to be due to this inevitable variability. There may be a number of ways of sizing up the position involving different assumptions or hypothetical constructs; he may follow one or several of these. The value of his advice is dependent almost

entirely on the soundness of his scientific judgement, and very little on whether his back-room calculations have been based on inverse or direct probability or on an appeal to fiducial argument.

11. How far, then, can one go in giving precision to a philosophy of statistical inference? It seems clear that in certain problems probability theory is of value because of its close relation to frequency of occurrence; such seems to be the case for my Example A. Tests can be built up to satisfy the practical requirements in this field. In other and, no doubt, more numerous cases there is no repetition of the same type of trial or experiment, but all the same we can and many of us do use the same test rules to guide our decision, following the analysis of an isolated set of numerical data. Why do we do this? What are the springs of decision? Is it because the formulation of the case in terms of hypothetical repetition helps to that clarity of view needed for sound judgement? Or is it because we are content that the application of a rule, now in this investigation, now in that, should result in a long-run frequency of errors in judgement which we control at a low figure? On this I should not care to dogmatize, realizing how difficult it is to analyse the reasons governing even one's own personal decisions.

12. That the frequency concept is not generally accepted in the interpretation of statistical tests is of course well known. With his characteristic forcefulness R. A. Fisher (1945*b*) has recently written: 'In recent times one often repeated exposition of the tests of significance, by J. Neyman, a writer not closely associated with the development of these tests, seems liable to lead mathematical readers astray, through laying down axiomatically, what is not agreed or generally true, that the level of significance must be equal to the frequency with which the hypothesis is rejected in repeated sampling of any fixed population allowed by hypothesis. This intrusive axiom, which is foreign to the reasoning on which the tests of significance were in fact based seems to be a real bar to progress....'

13. But the subject of criticism seems to me less an intrusive mathematical axiom than a mathematical formulation of a practical requirement which statisticians of many schools of thought have deliberately advanced. Prof. Fisher's contributions to the development of tests of significance have been outstanding, but such tests, if under another name, were discovered before his day and are being derived far and wide to meet new needs. To claim what seems to amount to patent rights over their interpretation can hardly be his serious intention. Many of us, as statisticians, fall into the all too easy habit of making authoritative statements as to how probability theory should be used as a guide to judgement, but ultimately it is likely that the method of application which finds greatest favour will be that which through its simplicity and directness appeals most to the common scientific user's understanding. Hitherto the user has been accustomed to accept the function of probability theory laid down by the mathematicians; but it would be good if he could take a larger share in formulating himself what are the practical requirements that the theory should satisfy in application.

(ii) The choice of statistical tests

14. One approach to follow in determining tests to be applied to the 2×2 class of problem follows the lines that Neyman and I have adopted since 1928 in dealing with tests of statistical hypotheses. Let me first recapitulate in broad terms the steps in that approach when applied to a problem where the universe of possible observations can be represented by a

finite set of discrete points. A test of significance may be described as a method of analysis of statistical data which helps us to discriminate between alternative theories or hypotheses. In order to make use of the theory of probability in the sense here understood, a random process must either have been purposely introduced or be assumed to have been present in the collection of data; then the hypothesis very often concerns the values of parameters contained in the probability laws which, in the conceptual sphere, form the mathematical counterpart of the sampling distributions of experience.

15. We proceed by setting up a specific hypothesis to test, H_0 in Neyman's and my terminology, the null hypothesis in R. A. Fisher's. At the same time, in choosing the test, we take into account alternatives to H_0 which we believe possible or at any rate consider it most important to be on the look out for. Thus we wish the test to have maximum discriminating power within a certain class of hypotheses. Three steps in constructing the test may be defined:

Step 1. We must first specify the set of results which could follow on repeated application of the random process used in the collection of the data; this may be termed the experimental probability set.

Step 2. We then divide this set by a system of ordered boundaries or contours such that as we pass across one boundary and proceed to the next, we come to a class of results which makes us more and more inclined, on the information available, to reject the hypothesis tested in favour of alternatives which differ from it by increasing amounts.

Step 3. We then, if possible, associate with each contour level the chance that, if H_0 is true, a result will occur in random sampling lying beyond that level.

This rather crude statement of procedure will be developed in more detail in discussing the problems that arise in connexion with the 2×2 table.

16. *Notes on these points.* (*a*) *Step* 1. This involves the definition of what Neyman and I have termed the sample space, W. The application in three forms of the 2×2 problem is discussed in paragraphs 19, 27 and 46 below.

(*b*) *Step* 2. For a given hypothesis under test there may be a number of ways of deriving a system of contours, and only in certain cases can there be said to be complete agreement on which is the 'best'. Practical expediency will often carry weight in the choice. It is widely accepted that the choice cannot be made without paying regard to the admissible hypotheses alternative to H_0, whether this process is given formal precision or taken as a broad guide. In our first papers (Neyman & Pearson, 1928*a*, *b*) we suggested that the likelihood ratio criterion, λ, was a very useful one to employ in determining a family of contours which would be ordered in relation to our confidence in the hypothesis tested when set against the background of admissible alternatives. Thus Step 2 preceded Step 3. In later papers (Neyman & Pearson, 1933, 1936 and 1938) we started with a fixed value for the chance, ϵ, of Step 3 and determined the associated contour, taking account of what we termed the power of a test with regard to the alternative hypotheses. The family of Step 2 followed on giving decreasing values to ϵ. However, although the mathematical procedure may put Step 3 before 2, we cannot put this into operation before we have decided, under Step 2, on the guiding principle to be used in choosing the contour system. That is why I have numbered the steps in this order.

(*c*) *Step* 3. If this can be accomplished, we have what Neyman and I called control of the '1st kind of error'. In problems where, as below, we are concerned with discrete rather than

continuous probability distributions (e.g. for the binomial, the Poisson, the multinomial and the hypergeometric distributions), this objective cannot always be achieved, and it may be necessary to be satisfied with a knowledge of an upper limit of the chance of rejecting the hypothesis tested when it is true.

(iii) Application of this approach to the analysis of data classed in a 2 × 2 table

17. The frequencies of the data in the table may be defined in the following notation:

Table 1

	Col. 1	Col. 2	Total
Row 1	a	c	m
Row 2	b	d	n
Total	r	s	N

If we follow in turn the steps defined above to determine the method of interpretation of such data, the requirements of the appropriate tests are seen to follow very simply, although mathematical or computational difficulties arise in implementing them. On taking Step 1 we can separate out at once the three types of problem which Barnard has differentiated;* these I shall call Problems I, II and III. They are distinguished by the sample space having 1, 2 and 3 dimensions respectively. From the mathematical point of view it might seem more logical to take them in the reverse order, adding first one and then a second restriction to the 3-dimensioned case of Problem III. For a simple exposition, I think the reverse procedure of building up from I to III is preferable and this has been adopted in the following sections.

(iv) Problem I

18. This may be described as the test of the significance of the difference between two treatments after these have been randomly assigned to a group of $N = m+n$ individuals (Barnard terms it the 2×2 independence trial). To use the terminology of a particular application, we may say that we are observing the presence or absence of 'reaction X'. The first treatment is applied to m and the second to n of the N individuals; as a result a/m and b/n show reaction X.

19. In this case the random process has been applied within the group of N individuals, and its repetition would simply involve other random reassignments of the two treatments among the N. No assumption is made as to how the N individuals were selected from some larger universe. The repetition may be hypothetical, in the sense that it often could not take place, e.g. if reaction X = death. Indeed, repetition under the same essential conditions is frequently impossible in practice. But this correspondence between the frequency of results upon hypothetical repetition and the probability distribution of the counterpart mathematical model forms an accepted part of the process of reasoning whereby (following

* Statisticians had, of course, all been more or less conscious of these differences, but, at any rate in my own case, it was discussion with Mr Barnard which made it easy to see the problem in its full clarity.

the present approach) we use probability theory as a basis for inference. The hypothesis tested is that while some individuals show reaction X and some do not, the result would be the same whichever treatment were applied *as far as these N individuals are concerned*. Thus, on the null hypothesis, there are $r = a+b$ individuals who will react and $s = c+d$ who will not, whatever the assignment of treatments.

20. The chance that a will react in m and $b = r-a$ in n is, therefore, if the hypothesis be true,

$$P_1\{a \mid N, r, m\} = \frac{m!\,n!\,r!\,s!}{a!\,b!\,c!\,d!\,N!}. \tag{1}$$

This expression is proportional to the coefficient of x^a in the hypergeometric series

$$F(\alpha, \beta, \gamma, x) = F(-r, -m, n-r+1, x). \tag{2}$$

Thus, taking $m \geqslant n$, a can assume values of

(i) $0, 1, \ldots, r$ if $r \leqslant n$,

(ii) $r-n, r-n+1, \ldots, r$ if $n < r \leqslant m$,

(iii) $r-n, r-n+1, \ldots, m$ if $r > m$.

For this probability distribution, it is known (K. Pearson (1899) and Kendall (1943, p. 127))

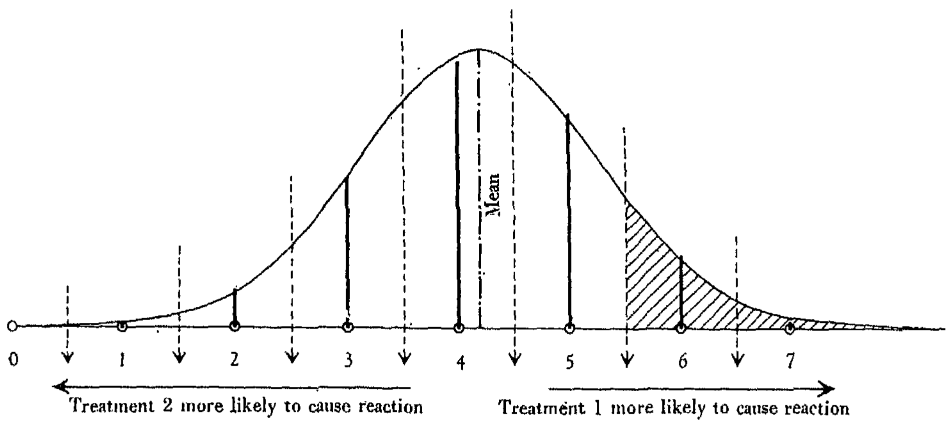


Fig. 2

that

$$\text{Mean}\, a = \frac{rm}{N}, \tag{3}$$

$$\text{Variance of } a = \sigma_a^2 = \frac{mnrs}{N^2(N-1)}. \tag{4}$$

21. For the particular case

$$N = 20, \quad r = 7, \quad m = 12, \quad n = 8,$$

the terms in the distribution of $P_1\{a \mid 20, 7, 12\}$ are shown as ordinates in Fig. 2 and given in the accompanying Table 2. The experimental probability set consists of the eight alternative values for a, viz. $0, 1, \ldots, 7$ with which the probabilities tabled are associated if H_0 is true. Further

$$\text{Mean}\, a = \bar{a} = 4\cdot 2, \quad \sigma_a = 1\cdot 0721. \tag{5}$$

22. Next consider step 2. The purpose of the investigation is to test the hypothesis that the difference between $a/12$ and $(r-a)/8$ has resulted simply from a random partition of 20 individuals, of whom r will show reaction X in whichever treatment group they are included. The experiment gives $r=7$. The contour levels fall between the 8 points of the set as shown in Fig. 2; the further a lies towards the right, the more *inclined we shall be to* accept the alternative hypothesis that $a/12 > (r-a)/8$ because treatment 1 is more effective than treatment 2. The further a lies to the left, the more we shall incline towards the reverse alternative. To complete Step 3, we have only to calculate the sums of the tail terms of the hypergeometric series, as shown in Table 2 for the special case.

Table 2. *Problem I. Chances for special case $N = 20$, $r = 7$, $m = 12$, if H_0 is true*

a	Chance of a	Chance of a or less	
		True value	Normal approx.
0	0·0001	0·000	0·000
1	0·0043	0·004	0·006
2	0·0477	0·052	0·056
3	0·1987	0·251	0·257
4	0·3576	—	—
		Chance of a or more	
		True value	Normal approx.
5	0·2861	0·392	0·390
6	0·0954	0·106	0·113
7	0·0102	0·010	0·016

23. Having set up the machinery of the test, we come to the practical question. Beyond which contour levels must a fall before we infer that there is a treatment difference? Not, I think, in the example, if a were 3, 4 or 5; possibly if $a = 6$, more probably if $a = 2$ and almost certainly if $a = 0$, 1 or 7. Were we to fix as critical levels those between $a = 1$ and 2 on the one hand, and between $a = 6$ and 7 on the other, then we should be guided in our decision by the following knowledge: if there were no treatment difference, so that seven out of the twenty individuals would have shown reaction X whichever treatment were applied, then the chance under random assignment of treatments that $a < 2$ or > 6 is only 0·014 or 1 in 70. Had we taken the critical levels between 2 and 3 and between 6 and 7, the corresponding chance would be 0·062 or 1 in 16. This summing up in terms of probability helps towards the balanced decision on the next practical step to be taken, because it helps us to assess the extent of purely chance fluctuations that are possible. It may be assumed that in a matter of importance we should never be content with a single experiment applied to twenty individuals; but the result of applying the statistical test with its answer in terms of the chance of a mistaken conclusion if a certain rule of inference were followed, will help to determine

the lines of further experimental work and the degree of confidence with which we proceed provisionally to adopt a new technique.

24. An experiment falling under this head has the advantage that the random process introduced is under complete control. The analysis will give an answer in probability terms whether the N individuals have been randomly selected from a larger whole or not. But this answer is limited in the sense that it relates only to the N; if we wish to draw conclusions about a wider population or populations, then a random selection of the N or, separately, of both its parts m and n is needed. Thus we come to Problems II and III.

25. *Approximation to the hypergeometric terms.* When dealing with small numbers, the calculation of the tail terms of the series may not be laborious, but it soon becomes so when r is large. An obvious approximation is that obtained by using an integral under the normal curve with the mean and standard deviation of equations (3) and (4) to represent the sum of the hypergeometric terms. As usual when approximating to the sum of the terms for $x = a, a+1, a+2, \ldots$, etc., of a discrete probability distribution by the integral under a continuous curve, we take this integral from the point $x = a - \frac{1}{2}$. Thus Fig. 3 shows the normal curve

$$p(x) = \frac{1}{\sqrt{(2\pi)}\,\sigma_a} \exp\left[-\tfrac{1}{2}(x-\bar{a})^2/\sigma_a^2\right], \tag{6}$$

with $\bar{a}$ and σ_a as in equations (5), and the approximation to the sum of the hypergeometric terms for $a = 6$ and 7 is

$$\int_{5.5}^{\infty} p(x)\,dx,$$

represented by the area marked with cross-hatching. The approximations for different levels are shown in Table 2, and are seen in this case to be quite adequate for the purpose of the test. Further comparisons are made in the Appendix, and it appears that provided m and n are fairly nearly equal, as they are likely to be in most planned experiments of the Problem I type, the normal approximation is surprisingly good. Yates (1934) has suggested a method of further correction.

26. *The correction for continuity.* In the 2×2 table connexion, the improvement obtained by taking the normal integral (i) from $x = a - \frac{1}{2}$ if $a > \bar{a}$ or (ii) from $x = a + \frac{1}{2}$ if $a < \bar{a}$ (so that we are summing for the lower tail), was pointed out by Yates (1934) and has often been termed 'Yates's correction for continuity'. It is, however, the natural adjustment to make on the basis of the Euler-Maclaurin theorem, when approximating to a sum of ordinates by an integral and without wishing to detract from the value of Yates's suggestion in this particular problem, it should be pointed out that the adjustment was used by statisticians well before 1934, when employing a normal or skew curve to give the sum of terms of a binomial or hypergeometric series.*

(v) PROBLEM II

27. This may be described as the test of whether the proportion of individuals bearing a character A is the same in two different populations, from each of which a random sample has been drawn, i.e. the test of the hypothesis that

$$p_1(A) = p_2(A) = p, \tag{7}$$

* The method was in use in the Department of Applied Statistics when I joined the staff in 1921, and may have been current many years before that.

where p is some common but unspecified proportion. Barnard describes this as the case of the 2×2 comparative trial. Here m individuals have been drawn at random from the first population and n from the second, and it is found that a/m and b/n, respectively, bear the character A. The conditions are assumed to be such that if the random procedure of selection were repeated, the appropriate probability distributions for a and b would be given by the terms of binomial expansions. Table 3 shows the observed results.

Table 3

	No. with character A	No. without A	Total
1st sample	a	c	m
2nd sample	b	d	n
Total	r	s	N

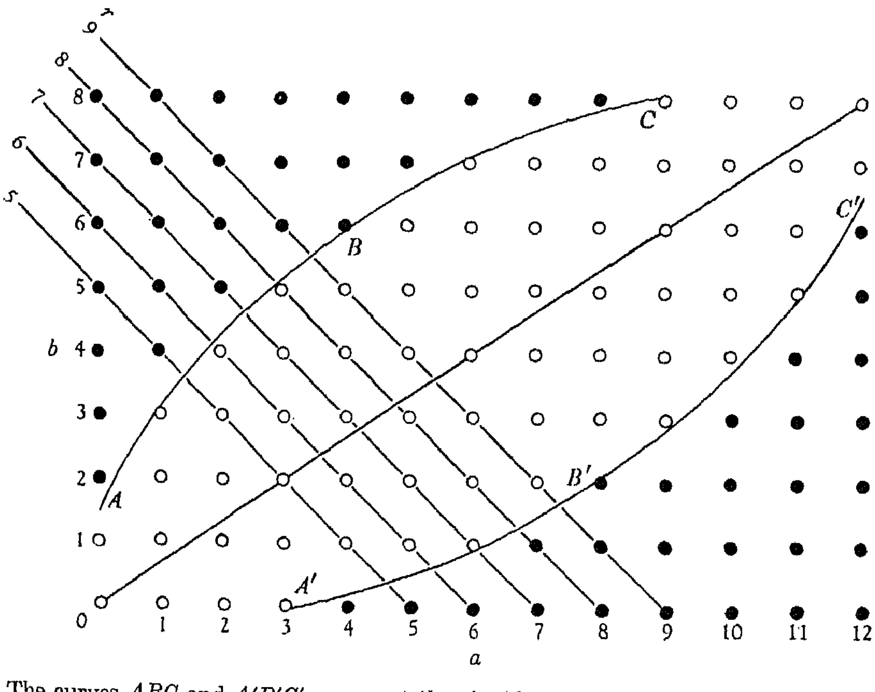


Fig. 3. The curves ABC and $A'B'C'$ represent the significance contours L_ϵ and L'_ϵ, respectively.

In this problem there have been two applications of a random selection process, not one as for Problem I, and the experimental probability set consists of the $(m+1)(n+1)$ alternative values of the doublet (a, b) $(0 \leqslant a \leqslant m, 0 \leqslant b \leqslant n)$ which can be represented in the lattice diagram shown in Fig. 3 for the special case $m = 12$, $n = 8$. It might, of course, be argued that in the hypothetical repetition of the selection process m and n need not remain constant, but this, I think, would introduce an unnecessary complication into the probability set-up.

28. The question before us is whether the result (a, b) is consistent with the hypothesis H_0 defined in equation (7) above, or whether it suggests that either $p_1 > p_2$ or that $p_1 < p_2$. A little reflexion shows that we have no reason to reject H_0 if the point (a, b) lies near the diagonal line on which $a/m = b/n$, but, broadly speaking, are more and more likely to do so the farther the point falls from this line in the direction of the corners $(0, n)$ and $(m, 0)$ of the lattice diagram. This statement requires amplification. In defining the significance contours we may consider the following question: If H_0 is not true, what departures from equality in p_1 and p_2 do we regard it of equal importance to detect? Should the power of the test be roughly the same for constant values, for example, of

$$(a)\quad p_1 - p_2, \qquad (b)\quad p_1/p_2 \qquad \text{or} \qquad (c)\quad \frac{p_1}{1-p_1}\bigg/\frac{p_2}{1-p_2}?$$

The procedure which I have adopted in the sections which follow is frankly one of expediency. I have not considered in detail how to choose a family of significance contours satisfying requirements formulated in advance, but have taken those suggested by the customary large-sample procedure which gives contours of the form ABC, $A'B'C'$ drawn in Fig. 3. These will, I believe, make the power of the test to detect a difference more nearly dependent on the ratio of the odds given by (c) than on either of the expressions (a) or (b). E. B. Wilson (1941) chooses the expression (a). This point, however, needs further investigation. It should be noted that a similar problem, in the case where the sampling distributions follow the Poisson law, was discussed very fully by Przyborowski & Wilenski (1939).

29. Besides involving a 2-dimensional instead of a 1-dimensional experimental probability set, Problem II differs from Problem I in that we need an answer which is independent of the unknown common probability p of the null hypothesis. In Problem I the part of p was played by the fraction r/N given by the data. We are concerned now with what Neyman and I (Neyman & Pearson, 1933) have termed a composite hypothesis, and were it possible would like the contour levels to bound regions which are 'similar to the sample space with regard to the parameter p' (loc. cit. p. 313) (i.e. are independent of p). The following considerations show the lines along which a first attack of the problem can proceed.

30. If H_0 is true and equation (7) holds, then the probability of the observed result may be written*

$$P_2\{a \mid p, m\} \times P_2\{b \mid p, n\} = \frac{m!}{a!\,c!}p^a(1-p)^c \times \frac{n!}{b!\,d!}p^b(1-p)^d \tag{8·1}$$

$$= \frac{N!}{r!\,s!}p^r(1-p)^s \times \frac{m!\,n!\,r!\,s!}{a!\,b!\,c!\,d!\,N!} \tag{8·2}$$

$$= P_2\{r \mid p, N\} \times P_1\{a \mid N, r, m\}. \tag{8·3}$$

Thus the probability of obtaining the doublet (a, b) in sampling from two populations with a common p may be regarded as the product of two terms:

(i) The probability that $a + b = r$ or that the point (a, b) in Fig. 3 falls on a diagonal line on which r = constant. This probability, $P_2\{r \mid p, N\}$, is the $(r+1)$th term in the expansion of the binomial

$$((1-p)+p)^N.$$

(ii) The relative probability, given r, of the observed partition into a and $b = r - a$; this is independent of p and is identical with the expression $P_1\{a \mid N, r, m\}$ of equation (1), i.e. is proportional to a term of the hypergeometric series (2).

* It will be seen that $P_1\{\ \}$ has been used to denote a hypergeometric probability and $P_2\{\ \}$ a binomial probability.

31. If, now, it were possible to draw a boundary line L_ϵ such as ABC shown in Fig. 3, cutting off at the end of each diagonal, r = constant, a group of points $(a, r-a)$ such that

$$\sum_a [P_1\{a \mid N, r, m\}] = \epsilon, \tag{9}$$

where ϵ is a fraction between 0 and 1 chosen at will, then the requirement of Step 3 would be satisfied. For in rejecting H_0 when (a, b) fall beyond this boundary,* the chance of doing so if H_0 were true would be

$$\sum_{r=0}^{N} [P_2\{r \mid p, N\} \times \epsilon] = \epsilon \times \sum_{r=0}^{N} [P_2\{r \mid p, N\}] = \epsilon, \tag{10}$$

i.e. would be independent of the unknown common p of the hypothesis tested. The test would then be analogous to 'Student's' test for the significance of the difference between two means, where we have a system of contour levels L_ϵ each associated with a chance ϵ, independent of the values of any unknown parameters which are irrelevant to the composite hypothesis tested.

32. Unfortunately, this objective cannot be achieved because we are not dealing with continuous probability distributions and $P_1\{a \mid N, r, m\}$ exists only at discrete, integral values of a. If we follow the present line of approach, all that is possible is to take contour or significance levels which cut off from an end of each diagonal, r = constant, a group of points for which

$$\sum_a [P_1\{a \mid N, r, m\}] = \beta_r \leqslant \epsilon. \tag{11}$$

Then, in rejecting H_0 when (a, b) falls beyond such a contour, we know that the chance of doing so, if H_0 is true, will be

$$\sum_{r=0}^{N} [P_2\{r \mid p, N\} \times \beta_r] \leqslant \epsilon. \tag{12}$$

It is clear that the amount by which the probability falls below ϵ will be a function of p, and that in taking Step 3 we are only associating with each significance level L_ϵ an upper limit, ϵ, to the probability of rejecting H_0 when it is true.

33. We have still, of course, to determine the most appropriate system of significance levels and to set out a ready means of finding an upper limit, ϵ, associated with the level on which an observed doublet (a, b) falls.† Mr Barnard has broken new ground in

(i) defining for this Problem II one systematic method of determining a family of levels L_ϵ based on certain clearly defined principles;

(ii) determining the true upper bound to the associated probability ϵ which, in the case of small samples at any rate, may be considerably below that which has hitherto been used.

Since, however, much tabling is needed before his theoretical advance can be followed by a practical working rule available for samples of any sizes, m and n, I think it is worth while describing the cruder handling of the lattice diagram which I had discussed in 1938–9

* There would be a similar series of boundaries, L'_ϵ, below the diagonal $a/m = b/n$, such as $A'B'C'$ of Fig. 3.

† The likelihood ratio λ might be used in determining the family of significance contours, as was suggested in connexion with the general χ^2 problem (Neyman & Pearson, 1928*b*, p. 283). In large samples λ would approximately equal $e^{-\frac{1}{2}u^2}$, where u is given by equation (22) below.

lectures. This involves, perhaps, not much more than a restatement of what may be termed the classical approach to Problem II (see paras. 43 and 44 below), but it does bring out the difference between Problems I and II, which I think important.

34. It may be well to emphasize here that this distinction between the handling of Problems I and II is not universally accepted. Fisher has set out his approach as follows in a paper read before the Royal Statistical Society (1935): 'To the many methods of treatment hitherto suggested for the 2×2 table the concept of ancillary information suggests this new one. Let us blot out the contents of the table, leaving only the marginal frequencies. If it be admitted that these marginal frequencies by themselves supply no information on the point at issue, namely, as to the proportionality of the frequencies in the body of the table, we may recognize the information they supply as wholly ancillary; and therefore recognize that we are concerned only with the relative probabilities of occurrence of the different ways in which the table can be filled in, subject to these marginal frequencies.'

This view has also been supported by Yates (1934). As I understand it, Fisher would refer the observation (a, b) to a linear set (as in my Problem I), however the data have been collected; this attitude follows readily if we discard the requirement that the probability distribution used in the test must be related to the frequency distribution that would be generated by repeated application of the random sampling process employed in the experiment. It will be seen that with Fisher's approach there is a gain in simplicity in handling the analysis; it must remain a matter of opinion whether there is a loss in the relevance of the probability construct to the question at issue. It is, of course, only when handling small samples or in cases where (a, b) lies close to one of the corners $(0, 0)$ or (m, n) of the lattice that this need for choice between probability constructs is thrust upon us.

(vi) Solution of Problem II, using the normal approximation

35. If the samples are large, the calculation of hypergeometric terms becomes laborious and we turn naturally, as in so many other statistical problems, to the approximation using the normal curve. In fact, except when r or s are very small or m and n very different in magnitude, the normal curve with mean and standard deviation given by equations (3) and (4) provides a surprisingly good approximation to the relative probability distribution of a for fixed r, viz. $P_1\{a \mid N, r, m\}$ (see Appendix). Define u_ϵ as the deviate of the standardized normal curve for which

$$\epsilon = \int_{u_\epsilon}^{\infty} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2} du \quad (\epsilon \leqslant \tfrac{1}{2}). \tag{13}$$

Then we can draw across the lattice diagram a significance level L_ϵ above and another L'_ϵ below* the diagonal $a/m = b/n$ such that

(i) all points (a, b) for which

$$\frac{(a+\frac{1}{2})-\bar{a}}{\sigma_a} \leqslant -u_\epsilon \tag{14}$$

lie beyond, i.e. above, L_ϵ;

(ii) and all points (a, b) for which

$$\frac{(a-\frac{1}{2})-\bar{a}}{\sigma_a} \geqslant u_\epsilon \tag{15}$$

lie beyond, i.e. below, L'_ϵ.

* The words 'above' and 'below' are used in the sense of Figs. 3 and 4.

If we wish to take special action either when a/m is significantly less than b/n or significantly greater, then we shall use both levels L_ϵ and L'_ϵ; if only, however, when $a/m < b/n$, then we use L_ϵ. The corresponding probability levels would be obtained by making ϵ for the second case twice its value for the first. Fig. 4 shows the 247 relative probabilities $P_1\{a \mid N, r, m\}$ for the case $m = 18$, $n = 12$. The unbroken, stepped lines are two contour levels determined in this way. Purely for convenience in drawing, the level with $\epsilon = 0{\cdot}05$ and $u_{0\cdot05} = 1{\cdot}6445$ has been put above the diagonal and that with $\epsilon = 0{\cdot}01$ and $u_{0\cdot01} = 2{\cdot}3263$ below.

36. If the normal approximation to the hypergeometric series were correct, it would follow that along every diagonal, r = constant, the sum of the relative probabilities for points above L_ϵ would satisfy the inequality (11). Hence the inequality (12) for the complete area of the lattice above L_ϵ would hold, whatever the value of the common p. A similar result would hold for the area below L'_ϵ. Of course, the normal approximation will not hold precisely, particularly when r or s are small, but here we shall generally be on the safe side, in the sense that the hypergeometric distribution is flat-topped with abrupt ends so that the β_r of equation (11) will be considerably less than ϵ, and often zero.

37. It is interesting to examine the results set out in Fig. 4 with the help of the detailed calculations given in Table 4. Columns (2) and (3) give, for constant r, the mean and standard deviation of $P_1\{a \mid 30, r, 18\}$, while columns (4) (for $L_{0\cdot05}$) and (8) (for $L'_{0\cdot01}$) give the cut-off points defined by the normal approximation, i.e.

$$a_1 = \bar{a} - \tfrac{1}{2} - u_{0\cdot05} \times \sigma_a \quad \text{and} \quad a_2 = \bar{a} + \tfrac{1}{2} + u_{0\cdot01} \times \sigma_a. \tag{16}$$

The sums of the relative probabilities $P_1\{a \mid 30, r, 18\}$ for $a \leqslant a_1$ and $a \geqslant a_2$ are given in cols. (5) and (9) respectively. Thus, for example, for $r = 7$

$$a_1 = 4{\cdot}2 - 0{\cdot}5 - 1{\cdot}6449 \times 1{\cdot}1543 = 1{\cdot}80,$$

and the sum of the probabilities for $a = 0$ and 1 is

$$0{\cdot}0004 + 0{\cdot}0082 = 0{\cdot}0086.$$

These are the tail sums, termed β_r in equation (11). It is clear from an examination of cols. (5) and (9) that they are all less, and many of them very much less than 0·05 and 0·01. This is inevitable with a discrete distribution containing few terms. The contour levels have been drawn conventionally in Fig. 4 as steps passing through the half-integer points and not through the cut-off points of cols. (4) and (8). Clearly, whichever way they are drawn, they will separate off the same subset of the $(m+1)(n+1)$ points in the lattice diagram

38. The next question is this. If we were to use either of these levels, what in fact would be the chance of the sample doublet (a, b) falling beyond, if the null hypothesis were true? This will depend on the common value of p. The product sums

$$\sum_{r=0}^{N} [P_2\{r \mid p, N\} \times \beta_r] = \sum_{r=0}^{N} \left[\frac{N!}{r!\,s!} p^r (1-p)^s \times \beta_r\right] \tag{17}$$

obtained by multiplying the expressions in cols. (5) and (9) of Table 4 by the appropriate binomial terms are shown for a variety of values of p in Table 5, cols. (2) and (3). It is clear at once how far on the safe side we are in saying that these chances are $\leqslant 0{\cdot}05$ and $0{\cdot}01$ respectively. Similar calculations were carried out for a second example, taking $m = n = 10$,

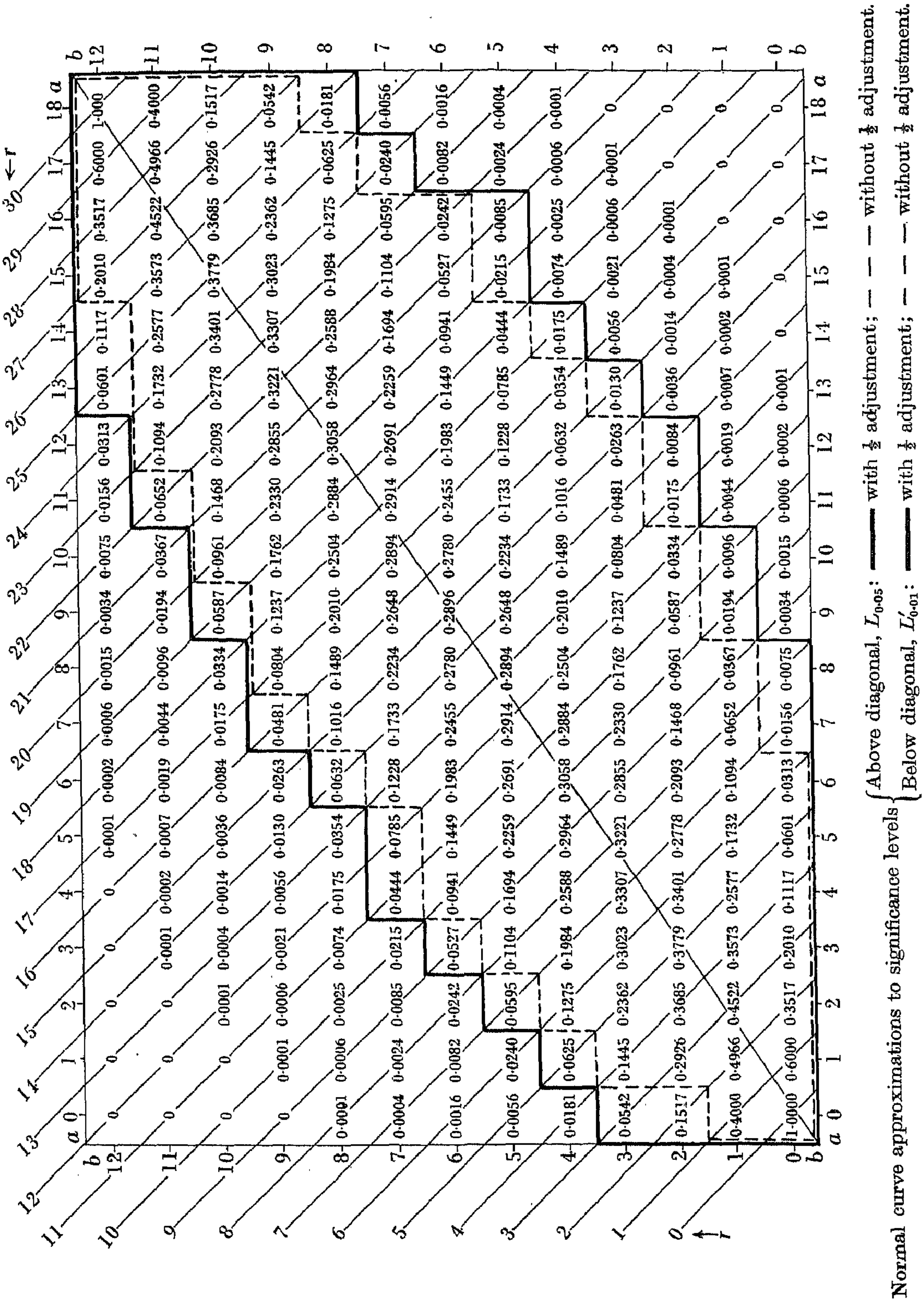


Fig. 4. Hypergeometric probabilities in lattice diagram for $m=18$, $n=12$.

Table 4. *Significance levels for case $m = 18$, $n = 12$*

r	$\bar{a}$	σ_a	Details for L_ϵ: $\epsilon = 0{\cdot}05$, $u_{0\cdot05} = 1{\cdot}6449$				Details for L'_ϵ: $\epsilon = 0{\cdot}01$, $u_{0\cdot01} = 2{\cdot}3263$				r
			Method 1		Method 2		Method 1		Method 2		
			Cut-off $\bar{a}-\frac{1}{2}-u_\epsilon\sigma_a$	Sum of terms beyond cut-off	Cut-off $\bar{a}-u_\epsilon\sigma_a$	Sum of terms beyond cut-off	Cut-off $\bar{a}+\frac{1}{2}+u_\epsilon\sigma_a$	Sum of terms beyond cut-off	Cut-off $\bar{a}+u_\epsilon\sigma_a$	Sum of terms beyond cut-off	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
0	0	0	−0·50	0	0	0	0·50	0	0	0	0
1	0·6	0·4899	−0·71	0	−0·21	0	2·24	0	1·74	0	1
2	1·2	0·6808	−0·42	0	0·08	0·1517	3·28	0	2·78	0	2
3	1·8	0·8187	−0·05	0	0·45	0·0542	4·20	0	3·70	0	3
4	2·4	0·9277	0·37	0·0181	0·87	0·0181	5·06	0	4·56	0	4
5	3·0	1·0171	0·83	0·0056	1·33	0·0681	5·87	0	5·37	0	5
6	3·6	1·0917	1·30	0·0256	1·80	0·0256	6·64	0	6·14	0	6
7	4·2	1·1543	1·80	0·0086	2·30	0·0681	7·39	0	6·89	0·0156	7
8	4·8	1·2069	2·31	0·0267	2·81	0·0267	8·11	0	7·61	0·0075	8
9	5·4	1·2507	2·84	0·0091	3·34	0·0618	8·81	0·0034	8·31	0·0034	9
10	6·0	1·2865	3·38	0·0241	3·88	0·0241	9·49	0·0015	8·99	0·0209	10
11	6·6	1·3152	3·94	0·0080	4·44	0·0524	10·16	0·0006	9·66	0·0102	11
12	7·2	1·3370	4·50	0·0197	5·00+	0·0982	10·81	0·0046	10·31	0·0046	12
13	7·8	1·3524	5·08	0·0414	5·58	0·0414	11·45	0·0020	10·95	0·0195	13
14	8·4	1·3615	5·66	0·0145	6·16	0·0777	12·07	0·0007	11·57	0·0091	14
15	9·0	1·3646	6·26	0·0301	6·76	0·0301	12·67	0·0038	12·17	0·0038	15
16	9·6	1·3615	6·86	0·0091	7·36	0·0572	13·27	0·0015	12·77	0·0145	16
17	10·2	1·3524	7·48	0·0195	7·98	0·0195	13·85	0·0060	13·35	0·0060	17
18	10·8	1·3370	8·10	0·0380	8·60	0·0380	14·41	0·0022	13·91	0·0197	18
19	11·4	1·3152	8·74	0·0102	9·24	0·0689	14·96	0·0080	14·46	0·0080	19
20	12·0	1·2865	9·38	0·0209	9·88	0·0209	15·49	0·0026	14·99	0·0241	20
21	12·6	1·2507	10·04	0·0401	10·54	0·0401	16·01	0·0006	15·51	0·0091	21
22	13·2	1·2069	10·71	0·0075	11·21	0·0727	16·51	0·0025	16·01	0·0025	22
23	13·8	1·1543	11·40	0·0156	11·90	0·0156	16·99	0·0086	16·49	0·0086	23
24	14·4	1·0917	12·10	0·0313	12·60	0·0313	17·44	0·0016	16·94	0·0256	24
25	15·0	1·0171	12·63	0	13·13	0·0601	17·87	0·0056	17·37	0·0056	25
26	15·6	0·9277	13·57	0	14·07	0·1117	18·26	0	17·76	0·0181	26
27	16·2	0·8187	14·35	0	14·85	0	18·60	0	18·10	0	27
28	16·8	0·6808	15·18	0	15·68	0	18·88	0	18·38	0	28
29	17·4	0·4899	16·09	0	16·59	0	19·04	0	18·54	0	29
30	18·0	0	17·50	0	18·00	0	18·50	0	18·00	0	30

and the results are shown in Table 5, cols. (6) and (7). In this case, the actual chances of (a, b) falling on or beyond the significance levels are even further below the nominal limits of 0·05 and 0·01. In fact, it becomes clear that in the case of small samples, at any rate, this method of introducing the normal approximation gives such an overestimate of the true chances of falling beyond a contour as to be almost valueless.

Table 5. *Showing the difference between nominal and actual significance levels*

p (if H_0 true)	1st example: $m = 18$, $n = 12$				2nd example: $m = 10 = n$				p (if H_0 true)
	Method 1		Method 2		Method 1		Method 2		
	True chance of falling on or beyond		True chance of falling on or beyond		True chance of falling on or beyond		True chance of falling on or beyond		
	$L_{0\cdot05}$	$L'_{0\cdot01}$	$L_{0\cdot05}$	$L'_{0\cdot01}$	$L_{0\cdot05}$	$L'_{0\cdot01}$	$L_{0\cdot05}$	$L'_{0\cdot01}$	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0·05	0·0010	0·0000	0·0478	0·0000	0·0000	0·0000	0·0069	0·0000	0·05
0·1	0·0054	0·0000	0·0602	0·0003	0·0005	0·0000	0·0251	0·0005	0·1
0·2	0·0141	0·0003	0·0483	0·0043	0·0037	0·0007	0·0455	0·0037	0·2
0·3	0·0174	0·0012	0·0490	0·0091	0·0058	0·0014	0·0495	0·0058	0·3
0·4	0·0204	0·0023	0·0542	0·0108	0·0062	0·0017	0·0546	0·0062	0·4
0·5	0·0219	0·0028	0·0498	0·0109	0·0062	0·0015	0·0572	0·0062	0·5
0·6	0·0221	0·0035	0·0437	0·0119	Repeat as for $1-p$				0·6
0·7	0·0204	0·0037	0·0431	0·0120					0·7
0·8	0·0126	0·0031	0·0459	0·0113					0·8
0·9	0·0019	0·0009	0·0282	0·0052					0·9
0·95	0·0001	0·0001	0·0058	0·0010					0·95

39. Before considering a second method, it will be useful to recapitulate certain characteristics of what I have termed Method 1. It provides for any nominal value of ϵ one systematic procedure of defining a critical boundary or significance level cutting off a region from the lattice diagram. Neither the subgroup of points cut off, nor the sum of the probabilities associated with them for a given p, will alter continuously with ϵ; they will change by discrete steps as the cut-off point, defined in para. 37, passes through a point (a, b). While we shall sometimes want to know whether the observed (a, b) falls beyond a level L_ϵ specified in advance, more often we shall ask what is the level on which (a, b) falls. This, using Method 1, we find by calculating

$$u = \frac{\bar{a} - (a + \frac{1}{2})}{\sigma_a} \quad \text{if} \quad a < \bar{a} \qquad \text{or} \qquad u = \frac{a - \frac{1}{2} - \bar{a}}{\sigma_a} \quad \text{if} \quad a > \bar{a}, \tag{18}$$

and finding ϵ from the normal integral of equation (13). In this way the nominal chance ϵ will be a little nearer the true upper limit than the figures in Table 5 suggest,* but not enough to modify the criticism expressed above.

* It will be seen from Table 4 that no point (a, b) gives a β_r in cols. (5) and (9) of exactly 0·05 or 0·01, respectively, so that no points actually lie on $L_{0\cdot05}$ or $L_{0\cdot01}$.

40. *Method* 2. The introduction of the correction of $\frac{1}{2}$ for continuity is certainly appropriate in using the normal approximation to the hypergeometric series in Problem I, but I think it is not helpful in Problem II where we are concerned with a 2-dimensional experimental probability set. If instead of obtaining significance levels L_ϵ and L'_ϵ as in paras. 35–37, we obtain them from inequalities similar to (14) and (15) but with the correction of $\frac{1}{2}$ omitted, then there are several points to be noted:

(*a*) For the significance level L_ϵ, the expression

$$\beta_r = \sum_a [P_1\{a \mid N, r, m\}], \tag{19}$$

where the summation is for values of a on the diagonal, r = constant, for which

$$a \leqslant a_1 = \bar{a} - u_\epsilon \times \sigma_a \tag{20}$$

will be sometimes less and sometimes greater than ϵ. Hence, in the balance, it seems likely that the chance of the point (a, b) lying beyond L_ϵ or

$$\sum_{r=0}^{N} \left[\frac{N!}{r!\,s!} p^r (1-p)^s \times \beta_r\right] \tag{21}$$

will lie closer to ϵ than when the $\frac{1}{2}$ correction is used. The position will be the same for L'_ϵ.

(*b*) In drawing repeated samples of m and n from two populations in which there is a common chance, p, of an individual possessing character A, the ratio

$$u = \frac{a - \bar{a}}{\sigma_a} = \frac{a - rm/N}{\sqrt{\dfrac{mnrs}{N^2(N-1)}}} \tag{22}$$

has, whatever be p, (i) an expectation of zero, (ii) a unit standard deviation.* The shape of the distribution will, of course, depend on p, but, *faut de mieux*, we may not in the long run do too badly by assuming it to be normal. It is, of course, the weighted combination of a number of hypergeometric series whose shape depends on r.

41. Consider the result of applying this Method 2 to the case $m = 18$, $n = 12$ already discussed. The procedure for determining the 0·05 and 0·01 significance levels will be exactly as under Method 1, except that the continuity correction of $\frac{1}{2}$ is omitted. The resulting levels are shown as dashed, stepped lines in Fig. 4.† They fall, on the whole, inside the significance levels obtained by Method 1. Now turn to Table 4, where cols. (6) and (10) show the cut-off points a half unit further in towards the diagonal $a/m = b/n$. Cols. (7) and (11) give the values of β_r; some of these are considerably above the nominal values of $\epsilon = 0{\cdot}05$ and $0{\cdot}01$, others are still well below. But from the approach to Problem II that has been adopted, this is immaterial since the experimental probability set is the 2-dimensioned one of the lattice diagram and is not restricted to the diagonal r = constant on which the observed point (a, b) may happen to lie. What we are concerned with is the summed chance given by expression (21) and the value of this is given for eleven values of p in cols. (4) and (5) of Table 5. It will be seen that this true chance does sometimes exceed the nominal values of 0·05 and 0·01,

* Provided cases where r or s are zero, making the expression (22) indeterminate with $u = 0/0$, are excluded. Mr Barnard has pointed out that one way of avoiding this exclusion would be to lay down that, when $u = 0/0$, we assign to the ratio a value chosen at random from a population (say normal) with zero mean and unit variance.

† Again, for convenience the 5 % level is drawn above and the 1 % level below the diagonal.

but never by very much. Again, for the second example with $m = 10 = n$ (Table 5, cols. (8) and (9)) the true chance, while it sometimes exceeds the nominal value, is always considerably nearer it than using the significance levels of Method 1.

42. It is clear that no final conclusions can be based on two numerical examples, but it seems that the test of the null hypothesis in Problem II should be carried out as follows:

(a) When m, n, r or s are small, with the help of tables prepared on Barnard's lines, based on an ordered classification of the points in the lattice diagram, and giving the true upper bound of the chance that a point (a, b) falls on or beyond the level on which the observed result lies. The particular basis of his classification may, of course, be modified.

(b) When m, n, r and s are large, by assuming that the u of equation (22) is a normal deviate with unit standard deviation.

(vii) THE CLASSICAL APPROACH TO PROBLEM II

43. It has recently become customary to regard the test of significance applied to data given in a 2×2 table as the limiting case of a χ^2 test with one degree of freedom. But Problem II was originally answered in somewhat different terms. It was noted that if

$$p_1(A) = p_2(A) = p, \tag{23}$$

then the fractions a/m and b/n would both have expectations of p and variances of $p(1-p)/m$ and $p(1-p)/n$, respectively. Hence, if the null hypothesis were true, the difference

$$d = \frac{a}{m} - \frac{b}{n} \tag{24}$$

would have

$$\left.\begin{aligned} \text{mean } d &= 0 \\ \sigma_d &= \sqrt{\left[p(1-p)\left(\frac{1}{m}+\frac{1}{n}\right)\right]} \end{aligned}\right\}. \tag{25}$$

In large samples, therefore, it might be expected that

$$\frac{d}{\sigma_d} = \frac{a/m - b/n}{\sqrt{[p(1-p)(1/m+1/n)]}} \tag{26}$$

would be approximately normally distributed. Since by the nature of the problem the common value of p was unknown, an estimate was made from the sample, namely,

$$\hat{p} = \frac{a+b}{m+n} = \frac{r}{N}. \tag{27}$$

Substituting this into equation (26), we have

$$\frac{d}{s_d} = \frac{a/m - b/n}{\sqrt{[(r/N)(1-r/N)(1/m+1/n)]}} \tag{28·1}$$

$$= \frac{a - rm/N}{\sqrt{\left(\frac{mnrs}{N^3}\right)}}. \tag{28·2}$$

44. The form (28·2) is easily derived from (28·1), if we remember that $b = r - a$, $s = N - r$ and $m + n = N$.* It is seen that the ratio d/s_d is identical with the ratio u of equation (22), except for a factor $\sqrt{[(N-1)/N]}$ which is unimportant in large samples. Thus the classical test is practically identical with that suggested in paras. 40–42 above, though the two tests are differently derived.

* A third alternative form is, of course, $(ad - bc)\sqrt{N}/\sqrt{(mnrs)}$.

(viii) Problem III

45. This may be described as the test for the independence of two characters A and B. It is supposed that the probability that an individual selected at random will possess character A is $p(A)$ and that he will not possess it is $p(\bar{A}) = 1-p(A)$. The corresponding probabilities for character B are $p(B)$ and $p(\bar{B}) = 1-p(B)$. Four alternative combinations of the characters may occur, which may be denoted by AB, $A\bar{B}$, $\bar{A}B$ and $\bar{A}\bar{B}$. The various probabilities are set out in Table 6A. If the null hypothesis, H_0, specifying the independence of A and B is true, then

$$p(AB) = p(A)\times p(B), \quad p(A\bar{B}) = p(A)p(\bar{B}), \quad \text{etc.} \tag{29}$$

To test the hypothesis, we have a random sample of N observations with frequencies of occurrence of the combinations AB, $A\bar{B}$, etc., which may be classified in the 2×2 scheme of Table 6B. The sampling conditions are such that the probabilities of Table 6A are the same for all individuals selected, or, in conventional terms, the sample is drawn from an infinite population. Barnard calls this problem that of the double dichotomy.

Table 6A. *Probabilities*

	A	$\bar{A}$	Total
B	$p(AB)$	$p(\bar{A}B)$	$p(B)$
$\bar{B}$	$p(A\bar{B})$	$p(\bar{A}\bar{B})$	$p(\bar{B})$
Total	$p(A)$	$p(\bar{A})$	1

Table 6B. *Sample data*

	A	$\bar{A}$	Total
B	a	c	m
$\bar{B}$	b	d	n
Total	r	s	N

46. In Problem III there is only one application of a random process, the selection of N individuals, each one of which must fall into one or other of four alternative categories. If the random process were repeated and another sample of N drawn, not only are the frequencies a, b, c and d free to vary, but also *both* marginal totals, i.e. m may change as well as r. The experimental probability set will therefore contain results (a,b,c,d) restricted by the conditions (i) that none of the frequencies can be negative and (ii) that

$$a+b+c+d = N. \tag{30}$$

Geometrically, as Barnard points out, the set can be represented in 3 dimensions by points at unit intervals within a tetrahedron obtained by placing on top of one another the series of 2-dimensioned lattices of dimensions

$$0\times n, \quad 1\times(n-1), \quad 2\times(n-2), \quad \ldots, \quad (m-1)\times 1, \quad m\times 0. \tag{31}$$

47. We are again testing a composite hypothesis and should like to determine a family of critical surfaces to be used as significance levels, dividing the points within the tetrahedron in such a way that the chance of the sample point (a,b,c,d)* lying outside a given surface L_ϵ is equal to ϵ, whatever the values of the unknown probabilities $p(A)$ and $p(B)$. But again, as in Problem II, owing to the discontinuity in the set of points, there are no 'similar

* In view of the condition (30), the point can be defined by three co-ordinates, e.g. as (a, b, c), (a, b, m) or (a, r, m). In view of the form of equation (32), the last system of co-ordinates will be used.

regions'. We note that if H_0 is true, the probability of the observed result is a term of the multinomial expansion, viz.

$$\frac{N!}{a!\,b!\,c!\,d!}p(AB)^a p(A\bar{B})^b p(\bar{A}B)^c p(\bar{A}\bar{B})^d$$

$$= \frac{N!}{a!\,b!\,c!\,d!}p(A)^{a+b}p(B)^{a+c}p(\bar{A})^{c+d}p(\bar{B})^{b+d}$$

$$= \frac{N!}{m!\,n!}p(B)^m(1-p(B))^n \times \frac{N!}{r!\,s!}p(A)^r(1-p(A))^s \times \frac{m!\,n!\,r!\,s!}{a!\,b!\,c!\,d!\,N!}$$

$$= P_2\{m \mid p(B), N\} \times P_2\{r \mid p(A), N\} \times P_1\{a \mid N, r, m\}. \tag{32}$$

Here, the notation of para. 30 has been repeated.

48. Thus the probability of obtaining a sample represented by the triplet (a, r, m) may be regarded, if the characters A and B are independent, as the product of three terms:

(i) The probability of drawing m individuals with character B in a random sample of N, i.e. the probability that (a, r, m) falls in a horizontal section of the tetrahedron on which m = constant. This is the $(m+1)$th term in the expansion of the binomial

$$\{(1-p(B))+p(B)\}^N.$$

(ii) The probability of drawing r individuals with character A in a random sample of N, i.e. the probability that (a, r, m) falls on the vertical section of the tetrahedron on which r = constant. This is the $(r+1)$th term in the expansion of

$$\{(1-p(A))+p(A)\}^N.$$

(iii) The probability, given m and r, of the observed partition within the 2×2 table. This term represents the relative probability associated with the points lying along a straight line m = constant, r = constant; it is, of course, the same expression as has arisen in Problems I and II and is proportional to a term in the hypergeometric series $F(-r, -m, n-r+1, 1)$.

49. We are faced with a situation similar to that met under Problem II. Were it possible to cut off from each line on which m = constant, r = constant, a group of points such that

$$\sum_a [P_1\{a \mid N, r, m\}] = \epsilon, \tag{33}$$

then the subset of points within the tetrahedron composed of the sum of these groups for all possible combinations of m and r would have the property required of a 'critical region' in a significance test: i.e. the chance that the point (a, r, m) is included in the region, if H_0 is true, would be ϵ whatever values the irrelevant probabilities $p(A)$ and $p(B)$ assumed. However, (33) cannot be satisfied in general, and all that is possible is to define a family of significance contours such that the chance of a sample point falling beyond any one of them, say L_ϵ, is $\leqslant \epsilon$. By using the normal approximation to the sum of the hypergeometric tail-terms with the correction for continuity as described in paras. 35–39 for Problem II, we shall be very much on the safe side, i.e. the formal level of ϵ is likely to be much above the true chance of falling beyond the level, whatever be $p(A)$ or $p(B)$. The presence of the two binomial terms in equation (32) instead of the single term in equation (8·3), makes it likely that the overestimation of ϵ will be greater in Problem III than in II. It is to be expected, therefore, that any any rate when neither m, n, r or s are too small, the better approximation will be obtained by referring the u of equation (22) to the normal probability scale.

50. The handling of Problem III is discussed briefly by Barnard on p. 136 above. There is clearly room for further investigation. The general nature of the approximation

involved is of course that which arises in every χ^2 test for goodness of fit or for independence in an $h \times k$ table, where we replace a distribution consisting of a finite set of probabilities at discrete points in multiple space by a continuous distribution for which integration outside ellipsoidal contours is straightforward.

(ix) General comment

51. The duties of the statistician lie at many levels. He may be required merely to apply an established technique of analysis to an assembly of numerical data and this application may result in a statement, based on probability theory, of a 'level of significance' or a 'confidence interval', which will be used by others. Or he may be called on to share in planning the investigation or experiment which is to provide the data and then to draw conclusions from their analysis which will lead to further action. In this final role he needs to bring into play faculties which are no monopoly of his calling, the qualities of sound judgement which are the characteristics of a well trained, scientific mind. In the weighing of evidence, the result of the statistical analysis, expressed in one or more conventional probability figures, is only one factor in the summing up; as important, may be, is the question of whether the mathematical model is a fair counterpart to the happenings in the observational field. In addition, there will often be much information coming from outside the range of the immediate investigation, yet hardly expressible in numerical terms, which must influence decision.

52. It is perhaps hard experience gained in certain fields of war-time research, where decisions had to be reached on statistical data far less ample than could be wished, which has forced my own attention to this question: What weight do we actually give to the precise value of a probability measure when reaching decisions of first importance? One subject for examination falling under this inquiry is clearly the logical basis of the reasoning process by which judgement is influenced as a result of the application of a test of significance. This was the theme on which this paper opened. The approach illustrated in the pages which followed is a personal one and is set down, with no claim to be the best, in order to provoke thought and discussion. There appears no short route to a right answer in this matter; each individual who hopes to use his own judgement to the full in drawing conclusions from the statistical analysis of sampling data, must decide for himself what he requires of probability theory.

53. In the approach which I have followed and illustrated on the analysis of data classed in a 2×2 table, the appropriate probability set-up is defined by the nature of the random process actually used in the collection of the data. Consideration of this point forms the initial step in the determination of the appropriate test. On this score, what I have termed Problems I, II and III are differentiated. The difference is fundamental and lies at the bottom of the dilemma to which the Barnard-Fisher correspondence in *Nature* drew attention. It can be illustrated on the following data, given in Table 7, where I shall suppose that the effect we are interested in is that making a significantly greater than b.

54. If (a) the results have been obtained by random assignment of Treatment 1 to eighteen out of thirty individuals and Treatment 2 to the remaining twelve, and

(b) we merely ask whether the results are consistent with the hypothesis that the treatments are equivalent as far as these thirty individuals are concerned, so that the difference between the proportions 15/18 and 5/12 may reasonably be ascribed to a chance fluctuation,

(c) we are then concerned with Problem I, i.e. simply with the probabilities associated with the points $(a, 20-a)$ on the diagonal $r = 20$ of Fig. 4. The chance of getting $a \geqslant 15$, if the null hypothesis is true, is 0·0241,* or, using a common phrase, we can speak of the result being significant at the 2·5 % level.

55. On the other hand, if a sample of 18 has been drawn randomly from one population and a sample of 12 independently from a second and we wish to test whether $p_1(A) = p_2(A)$, then it seems to be an artificial procedure to restrict the experimental probability set to the 11 points on the line $r = 20$, i.e. to the values of a: 8, 9, ..., 18. A repetition of the double sampling process could give us a result (a, b) falling at any of the $19 \times 13 = 247$ points in the lattice diagram of Fig. 4. There will be a number of ways of defining a family of significance levels for this 2-dimensioned set; if we adopt that discussed in paras. 40–41, which

Table 7

For problem I	For problem II	Frequency of results A	Frequency of results $\bar{A}$	Total
1st treatment	Sample from 1st population	$a = 15$	$c = 3$	$m = 18$
2nd treatment	Sample from 2nd population	$b = 5$	$d = 7$	$n = 12$
	Total	$r = 20$	$s = 10$	$N = 30$

gives as two of its members the dotted, stepped lines shown in Fig. 4, we can say that the chance of a result falling beyond the lower line is certainly less than 0·015.† The observed point, with $a = 15$, $b = 5$ falls beyond the line, so that the result is undoubtedly 'significant at the 1·5 % level'.

56. These two probabilities, 2·5 and 1·5 %, are not the same, but there is no inconsistency in their difference. The character of the two investigations is different and to treat Problem II as though it were Problem I seems to call for a probability set-up which is unnecessarily artificial, when a simpler one is available. Admittedly by getting what seems to me a closer relation between the probability set-up and the experimental procedure, we have sacrificed some simplicity in handling the 2×2 table. But this is only the case when dealing with small numbers. For large numbers the methods of handling Problems I, II and III become, practically, identical.

57. Consider again the heavy shell problem described in para. 7 above. If we are to introduce probability theory, it seems to me that we should regard the problem as one in which we have a sample of $m = 12$ from the possible output of shell made to one design or by one firm and of $n = 8$ from the possible output of a second. This sampling may be hypothetical in that these may be 'pilot' shell, the first off production; nevertheless, this construct is

* For the normal curve approximation, using the correction for continuity, we find
$$u = (15 - \tfrac{1}{2} - 12{\cdot}0)/1{\cdot}2865 = 1{\cdot}943.$$
The proportionate area under the normal curve beyond this deviation is 0·026.

† Table 5, col. (5) shows the largest value of this chance to be 0·0120 for $p = 0{\cdot}3$. This figure cannot be much exceeded for other p's though I have not determined the precise maximum. I give 0·015 as a safe-side limit.

clearly less artificial than one in which, on the null hypothesis, we regard the experiment as though it were made on twenty shells, to twelve of which has been randomly assigned the label 'Made by firm X' and to the other eight, 'Made by firm Y'.

58. It is clear that in the heavy shell problem there may be many reasons to doubt whether the rounds fired can be regarded as a random sample from future output. That is why I have emphasized that the exploration which the statistician makes in private will not necessarily be presented in figures at the conference table. In this example, the proportions of successful perforations were 2/12 and 5/8; these put us on the line, $r = 7$, of the lattice diagram for which the hypergeometric probabilities were shown in Fig. 2. The sum of the terms with $a \leqslant 2$ is 5·2 % (normal approximation, using the $\frac{1}{2}$-correction, 5·6 %). This is the chance of getting as great or a greater positive difference, $b-a$, if H_0 were true, treating the case as Problem I. Barnard's method has not yet been extended to cover this case, but if we were to use the large sample method for handling Problem II, described in my paras. 40–41, we should find from equation (22) that

$$u = (2-4{\cdot}2)/1{\cdot}072 = -2{\cdot}05,$$

which puts (a, b) outside the upper 2·5 % level.

59. Were the action taken to be decided automatically by the side of the 5 % level on which the observation point fell, it is clear that the method of analysis used would here be of vital importance. But no responsible statistician, faced with an investigation of this character, would follow an automatic probability rule. The result of either approach would raise considerable doubts as to whether the performance of the first type of shell was as good as that of the second, but without the whole background of the investigation it is impossible to say what the statistician's recommendation as to further action would be.

60. In the example of the proof of anti-tank shot discussed in para. 6, the chance of perforation, p, while varying from plate to plate and batch to batch, will almost certainly not range through the whole interval 0–1. The striking-velocity of the shot would also probably be adjusted so that for average proof-plate and batches, p was near $\frac{1}{2}$. Then the discriminating level (or levels*) set across the 13 × 13 lattice diagram would be fixed paying regard to the likely variation in p; thus a fairly close upper limit could be calculated to the true probability of (a, b) falling beyond the level if the fresh batch were of the same quality as the standard. This is the upper limit of the risk of segregating the batch wrongly.

61. Precisely similar problems arise for consideration in even more difficult form in the analysis of data arranged in a $h \times k$ table, where h or k or both are > 2. It has become common practice to speak of the solution of this problem in terms of 'fixed marginal totals', but it may be questioned whether the restriction in the experimental probability set implied is generally appropriate. The frequencies in a $h \times k$ table may have been obtained by many different sampling procedures for, as in the 2 × 2 problem, a single form of tabular presentation will follow from a variety of types of investigation. For most of these, a repetition of the random process of selection would give results with either one or both sets of marginal totals changed.

62. For convenience in solution we may, of course, start by considering the distribution of our test criterion, on the null hypothesis, within the sub-set of results for which the margins

* It is possible that two levels might be taken with the associated proof rules: (i) if (a, b) falls beyond the outer one, reject the batch; (ii) if between outer and inner, fire further rounds; (iii) if within the inner level, accept the batch.

are fixed. If this distribution were the same whatever these fixed values, then the overall, distribution for unrestricted sampling would be the same as that for variation subject to fixed margins. Thus, mathematically, the solution of the partial problem would be a step in the solution of the complete one. But when applying χ^2 analysis to an $h \times k$ table, this result is only true as a large-sample approximation.

63. If we use the mathematical model which it is suggested gives the most direct aid in reasoning from the observations, i.e. that which regards the experimental probability set as generated by a repetition of the random process of selection used in collecting the data, then in the majority of cases we cannot regard the marginal totals as fixed. Thus a rigorous treatment would lead, as in the case of the 2×2 table, to a differentiation into a number of solutions. It is to be hoped, however,* unless the numbers in the margins are very small, that the χ^2 approximation with its appropriate degrees of freedom† will give results which are not misleading. This approximation leads, of course, in the 2×2 table to the reference of the ratio u of equation (22) to the normal probability scale. Some aspects of the approximation in this more general case were discussed by Yates (1934, pp. 233–35).

64. In closing I should like again to acknowledge my indebtedness to Mr G. A. Barnard. Having had the good fortune to discuss these problems with him and see drafts of his work over a period of 2 or 3 years it is difficult to say how many of his ideas have been built unconsciously into my own earlier approach. But I am especially aware of the clarification which his emphasis on the distinction between Problems I, II and III brought to my survey. I am also very grateful to Mr M. G. Kendall, Dr R. C. Geary and Dr B. L. Welch for a number of helpful criticisms, and to Mrs Maxine Merrington for her extensive computing work, which has alone made possible the various numerical illustrations that I have given.

* From the point of view both of the exponents of the fixed marginal and unrestricted marginal approach.

† The statement that, for example, in applying the test of independence of two characters to an $h \times k$ table, the degrees of freedom are $(h-1) \times (k-1)$, does not of course mean that sampling is restricted by fixed marginal totals. All that is implied is that approximately the overall distribution of the χ^2 function of the observations used, is the same as that for sampling within the restricted sub-set; this is because the distribution within each sub-set is approximately independent of the particular marginal totals which define it.

REFERENCES

BARNARD, G. A. (1945*a*). *Nature, Lond.*, **156**, 177.
BARNARD, G. A. (1945*b*). *Nature, Lond.*, **156**, 783.
FISHER, R. A. (1935). *J. Roy. Statist. Soc.* **98**, 39.
FISHER, R. A. (1941). *Science*, **94**, 210.
FISHER, R. A. (1945*a*). *Nature, Lond.*, **156**, 388.
FISHER, R. A. (1945*b*). *Sankhyā*, **7**, 130.
KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**. London: Charles Griffin and Co. Ltd.
NEYMAN, J. & PEARSON, E. S. (1928*a*). *Biometrika*, **20 A**, 195.
NEYMAN, J. & PEARSON, E. S. (1928*b*). *Biometrika*, **20 A**, 263.
NEYMAN, J. & PEARSON, E. S. (1933). *Philos. Trans.* A, **231**, 289.
NEYMAN, J. & PEARSON, E. S. (1936). *Statist. Res. Mem.* **1**, 113.
NEYMAN, J. & PEARSON, E. S. (1938). *Statist. Res. Mem.* **2**, 25.
PEARSON, K. (1899). *Phil. Mag.* **47**, 236.
PRZYBOROWSKI, J. & WILENSKI, H. (1939). *Biometrika*, **13**, 313.
WILSON, E. B. (1941). *Science*, **93**, 557.
WILSON, E. B. (1942). *Proc. Nat. Acad. Sci., Wash.*, **28**, 94.
YATES, F. (1934). *J. Roy. Statist. Soc.* Suppl. **1**, 217.

APPENDIX

THE NORMAL CURVE APPROXIMATION IN PROBLEM I

1. The following Tables 8 and 9 (A), (B) and (C) show the order of accuracy which results from using the normal curve integral as an approximation to the tail sums in the series

$$P_1\{a \mid N, r, m\} = \frac{m!\,n!\,r!\,s!}{a!\,b!\,c!\,d!\,N!} \tag{34}$$

the terms of which are proportional to those in the hypergeometric series

$$F(-r, -m, N-m-r+1, 1).$$

Here a is a variable which can assume the range of positive, integral values indicated under (i), (ii) and (iii) in para. 20 above, while N, r and m are fixed. The relation between these quantities and b, c, d, n and s is given in Table 1, para. 17. The method of approximation, using the '$\frac{1}{2}$' correction for continuity, has been discussed in para. 25.

2. Table 8 takes the case of an equal partition, $m = n = \frac{1}{2}N$, and shows the sum of the terms in the expression (34) for which $a \geqslant a_1$ which is also the sum of terms for which $a \leqslant r - a_1$. For $m \neq n$, results are given in Table 9 for $m > n$ and for the following proportionate partitions of N:

(A) $m = \frac{3}{5}N$, $n = \frac{2}{5}N$; (B) $m = \frac{4}{5}N$, $n = \frac{1}{5}N$; (C) $m = \frac{9}{10}N$, $n = \frac{1}{10}N$.

Here sums of terms at both tails of the series are needed. The sums (or chances of $a \geqslant a_1$ or $\leqslant a_1$) have not been given for all possible values of a_1 but, broadly speaking, for those within the limits where significance is likely to be in question. Sums below 0·0010 have generally been omitted. In each case the true sum of the terms (34) is compared with the approximation from the normal integral.

3. In drawing conclusions from the comparison, we have to decide what degree of accuracy is called for. Clearly the normal integral does not give mathematically exact results to 4 decimal places. On the other hand, except for certain instances where the partition is very unequal ($m = \frac{4}{5}N$ and $\frac{9}{10}N$) and r is small, the order of the approximation may be said to follow that of the series closely. If decisions are made by rule of thumb, according to the side of the 5 % or 1 % significance level on which a falls, then there are a number of entries in the tables where the approximation would give a on the wrong side. But one may question whether judgement of significance based on a single experiment can in fact be made sensitive to a difference between, say, 0·06 and 0·04 (odds of 16 to 1 and 24 to 1) or between 0·012 and 0·008 (odds of 82 to 1 and 124 to 1) and, given such latitude in accuracy, the approximation will be found generally sufficient. These must be points, however, where personal opinions will differ. Whatever views are held, the tables are sufficiently extensive to make it possible to obtain from them a rough measure of the accuracy of approximation in a wide range of cases.

4. It will be noted that in the symmetrical case ($m = \frac{1}{2}N$) and also when $m = \frac{3}{5}N$ the normal approximation for the tail sum is almost invariably a little too large. Undoubtedly for the symmetrical case an improved approximation could be obtained by modifying the $\frac{1}{2}$ correction used in calculating the ratio of deviation to standard deviation. This second order term would, however, need to vary with the probability level, thus complicating the procedure.

Table 8. *Case of equal partition, $m = n = \frac{1}{2}N$. Chance that $a \geqslant a_1$ = chance that $a \leqslant r - a_1$*

Partition		$m=n=50$		$m=n=30$		$m=n=20$		$m=n=15$		$m=n=10$			
r	a_1	True	Normal approx.	True	Normal approx.	True	Normal approx.	True	Normal approx.	True	Normal approx.	a_1	r
30	17	0·2566	0·2574	0·2194	0·2212							17	30
	18	·1376	·1388	·0981	·1002							18	
	19	·0630	·0643	·0348	·0365							19	
	20	·0243	·0253	·0096	·0106							20	
	21	·0078	·0085	·0020	·0024							21	
	22	·0021	·0024									22	
20	12	0·2269	0·2278	0·2060	0·2076	0·1715	0·1745					12	20
	13	·1053	·1068	·0852	·0873	·0564	·0592					13	
	14	·0392	·0408	·0270	·0287	·0128	·0144					14	
	15	·0114	·0126	·0064	·0073	·0019	·0025					15	
	16	·0025	·0031	·0011	·0014							16	
15	9	0·2884	0·2887	0·2760	0·2772	0·2572	0·2595	0·2330	0·2364			9	15
	10	·1312	·1325	·1163	·1185	·0954	·0985	·0715	·0755			10	
	11	·0453	·0473	·0358	·0380	·0242	·0265	·0134	·0156			11	
	12	·0113	·0129	·0077	·0090	·0040	·0049	·0014	·0020			12	
	13	·0019	·0027	·0011	·0016							13	
10	7	0·1589	0·1599	0·1495	0·1514	0·1367	0·1397	0·1226	0·1266	0·0894	0·0955	7	10
	8	·0458	·0486	·0399	·0429	·0324	·0357	·0251	·0285	·0115	·0147	8	
	9	·0078	·0101	·0061	·0081	·0042	·0058	·0026	·0038	·0005	·0011	9	
	10	·0006	·0014	·0004	·0010							10	
7	5	0·2179	0·2177	0·2119	0·2126	0·2038	0·2056	0·1950	0·1980	0·1749	0·1804	5	7
	6	·0558	·0594	·0514	·0553	·0458	·0501	·0401	·0448	·0286	·0338	6	
	7	·0062	·0096	·0053	·0084	·0042	·0068	·0032	·0055	·0015	·0031	7	
5	4	0·1810	0·1806	0·1766	0·1771	0·1709	0·1735	0·1648	0·1677	0·1517	0·1571	4	5
	5	·0281	·0339	·0261	·0320	·0236	·0295	·0211	·0270	·0163	·0220	5	

Table 9. *Case of unequal partition. Chances that $a \leqslant a_1$ and $a \geqslant a_1$*

(A) $m = \frac{3}{5}N$, $n = \frac{2}{5}N$

Partition			$m = 60$, $n = 40$		$m = 36$, $n = 24$		$m = 24$, $n = 16$		$m = 18$, $n = 12$		$m = 12$, $n = 8$				
r	Chance that	a_1	True	Normal approx.	True	Normal approx.	True	Normal approx.	True	Normal approx.	True	Normal approx.	a_1	Chance that	r
30	$a \leqslant a_1$	11	0·0020	0·0019									11	$a \leqslant a_1$	30
		12	·0074	·0074	0·0016	0·0020							12		
		13	·0230	0231	·0084	·0093	—						13		
		14	·0601	·0604	·0320	·0337	0·0023	0·0050					14		
		15	·1330	·1339	·0936	·0957	·0270	·0329					15		
		16	·2512	·2531	·2148	·2165	·1311	·1348					16		
	$a \geqslant a_1$	20	0·2533	0·2531	0·2148	0·2165	0·1322	0·1348					20	$a \geqslant a_1$	
		21	·1323	·1339	·0936	·0957	·0318	·0329					21		
		22	·0580	·0604	·0320	·0337	·0045	·0050					22		
		23	·0209	·0231	·0084	·0093							23		
		24	·0061	·0074	·0016	·0020							24		
		25	·0014	·0019									25		
20	$a \leqslant a_1$	6	0·0027	0·0026	0·0010	0·0012							6	$a \leqslant a_1$	20
		7	·0114	·0112	·0060	·0063	0·0015	0·0021	—				7		
		8	·0381	·0378	·0255	·0262	·0112	·0128	0·0015	0·0033			8		
		9	·1019	·1021	·0816	·0829	·0526	·0555	·0290	·0260			9		
		10	·2211	·2232	·2005	·2028	·1665	·1695	·1170	·1218			10		
	$a \geqslant a_1$	14	0·2236	0·2232	0·2017	0·2028	0·1665	0·1695	0·1182	0·1218			14	$a \geqslant a_1$	
		15	·0994	·1021	·0798	·0829	·0526	·0555	·0241	·0260			15		
		16	·0341	·0378	·0233	·0262	·0112	·0128	·0026	·0033			16		
		17	·0086	·0112	·0048	·0063	·0015	·0021					17		
		18	·0015	·0026	·0006	·0012							18		
15	$a \leqslant a_1$	4	0·0053	0·0053	0·0032	0·0033	0·0013	0·0015					4	$a \leqslant a_1$	15
		5	·0236	·0233	·0171	·0173	·0098	·0106	0·0038	0·0052			5		
		6	·0776	·0775	·0650	·0657	·0481	·0499	·0301	·0335	—		6		
		7	·1948	·1968	·1804	·1827	·1588	·1618	·1317	·1358	0·0511	0·0616	7		
	$a \geqslant a_1$	11	0·1970	0·1968	0·1814	0·1827	0·1587	0·1618	0·1317	0·1358	0·0578	0·0616	11	$a \geqslant a_1$	
		12	·0734	·0775	·0614	·0657	·0458	·0499	·0301	·0335	·0036	·0051	12		
		13	·0188	·0233	·0138	·0173	·0082	·0106	·0038	·0052	—		13		
		14	·0029	·0053	·0018	·0033	·0008	·0015					14		
10	$a \leqslant a_1$	2	0·0088	0·0089	0·0067	0·0071	0·0045	0·0050	0·0026	0·0033	0·0004	0·0009	2	$a \leqslant a_1$	10
		3	·0457	·0453	·0395	·0398	·0318	·0329	·0241	·0260	·0099	·0131	3		
		4	·1538	·1549	·1447	·1464	·1322	·1348	·1182	·1218	·0849	·0910	4		
	$a \geqslant a_1$	8	0·1539	0·1549	0·1442	0·1464	0·1311	0·1348	0·1170	0·1218	0·0849	0·0910	8	$a \geqslant a_1$	
		9	·0386	·0453	·0334	·0398	·0270	·0329	·0209	·0260	·0099	·0131	9		
		10	·0044	·0089	·0034	·0071	·0023	·0050	·0015	·0033	·0004	·0009	10		
7	$a \leqslant a_1$	0	0·0012	0·0022	0·0009	0·0013	0·0006	0·0010	0·0004	0·0007			0	$a \leqslant a_1$	7
		1	·0156	·0189	·0134	·0140	·0109	·0118	·0086	·0097	0·0044	0·0059	1		
		2	·0884	·0956	·0827	·0832	·0756	·0770	·0681	·0704	·0521	·0564	2		
	$a \geqslant a_1$	6	0·1492	0·1587	0·1426	0·1450	0·1341	0·1378	0·1250	0·1300	0·1056	0·1127	6	$a \geqslant a_1$	
		7	·0241	·0385	·0216	·0306	·0186	·0269	·0156	·0232	·0102	0·0160	7		
5	$a \leqslant a_1$	0	0·0088	0·0099	0·0078	0·0090	0·0066	0·0080	0·0056	0·0070	0·0036	0·0051	0	$a \leqslant a_1$	5
		1	·0816	·0811	·0778	·0781	·0730	·0742	·0681	·0701	·0578	·0616	1		
	$a \geqslant a_1$	5	0·0725	0·0811	0·0690	0·0781	0·0646	0·0742	0·0601	0·0701	0·0511	0·0616	5	$a \geqslant a_1$	

Table 9 (*continued*)

(B) $m = \frac{4}{5}N$, $n = \frac{1}{5}N$

Partition			$m = 80$, $n = 20$		$m = 48$, $n = 12$		$m = 32$, $n = 8$	
r	Chance that	a_1	True	Normal approx.	True	Normal approx.	True	Normal approx.
30	$a \leqslant a_1$	18	0·0018	0·0014				
		19	·0084	·0073	0·0013	0·0020		
		20	·0306	·0288	·0106	·0125		
		21	·0884	·0874	·0521	·0548		
		22	·2046	·2078	·1667	·1685		
	$a \geqslant a_1$	26	0·2092	0·2078	0·1667	0·1685		
		27	·0824	·0874	·0521	·0548		
		28	·0227	·0288	·0106	·0125		
		29	·0039	·0073	·0013	·0020		
		30	·0003	·0014				
20	$a \leqslant a_1$	11	0·0040	0·0026	0·0013	0·0011	—	
		12	·0182	·0148	·0095	·0087	0·0016	0·0031
		13	·0638	·0600	·0460	·0448	·0218	·0255
		14	·1729	·1755	·1523	·1542	·1176	·1208
	$a \geqslant a_1$	18	0·1758	0·1755	0·1522	0·1542	0·1176	0·1208
		19	·0499	·0600	·0371	·0448	·0218	·0255
		20	·0066	·0148	·0041	·0087	·0016	·0031
15	$a \leqslant a_1$	7	0·0018	0·0009	0·0008	0·0004		
		8	·0107	·0074	·0064	·0049	0·0022	0·0024
		9	·0462	·0408	·0355	·0323	·0217	·0219
		10	·1470	·1480	·1329	·1338	·1115	·1133
	$a \geqslant a_1$	14	0·1453	0·1480	0·1294	0·1338	0·1079	0·1133
		15	·0262	·0408	·0206	·0323	·0141	·0219
10	$a \leqslant a_1$	4	0·0039	0·0019	0·0026	0·0013	0·0012	0·0008
		5	·0254	·0191	·0206	·0159	·0145	·0121
		6	·1095	·1068	·1012	·0988	·0893	·0882
	$a \geqslant a_1$	10	0·0951	0·1068	0·0868	0·0988	0·0761	0·0882
7	$a \leqslant a_1$	2	0·0033	0·0013	0·0024	0·0010	0·0015	0·0007
		3	·0282	·0203	·0246	·0181	·0201	·0155
		4	·1408	·1417	·1354	·1364	·1281	·1293
	$a \geqslant a_1$	7	0·1985	0·1910	0·1906	0·1848	0·1805	0·1776
5	$a \leqslant a_1$	1	0·0053	0·0022	0·0045	0·0021	0·0035	0·0016
		2	·0531	·0434	·0499	·0430	·0457	·0383
	$a \geqslant a_1$	5	0·3193	0·2841	0·3135	0·2835	0·3060	0·2776

(C) $m = \frac{9}{10}N$, $n = \frac{1}{10}N$

Partition			$m = 90$, $n = 10$	
r	Chance that	a_1	True	Normal approx.
30	$a \leqslant a_1$	22	0·0009	0·0006
		23	·0073	·0057
		24	·0388	·0352
		25	·1384	·1388
	$a \geqslant a_1$	29	0·1356	0·1388
		30	·0229	·0352
20	$a \leqslant a_1$	14	0·0039	0·0019
		15	·0254	·0191
		16	·1095	·1068
	$a \geqslant a_1$	20	0·0951	0·1068
15	$a \leqslant a_1$	9	0·0006	0·0001
		10	·0063	·0027
		11	·0408	·0316
		12	·1705	·1765
	$a \geqslant a_1$	15	0·1808	0·1765
10	$a \leqslant a_1$	5	0·0006	0·0001
		6	·0082	·0029
		7	·0600	·0486
		8	·2615	·2902
	$a \geqslant a_1$	10	0·3305	0·2902
7	$a \leqslant a_1$	3	0·0016	0·0003
		4	·0207	·0096
		5	·1442	·1492
	$a \geqslant a_1$	7	0·4667	0·3974
5	$a \leqslant a_1$	2	0·0067	0·0006
		3	·0769	·0538
	$a \geqslant a_1$	5	0·4163	0·5000

2×2 TABLES. A NOTE ON E. S. PEARSON'S PAPER

By G. A. BARNARD

As Prof. Pearson has kindly shown me the proof of his paper, I should like to make the following further remarks.

1. If we have a sample of N from a population in which there is a chance p that an individual will have a character A, we can represent it in the form

$$x_1, x_2, \ldots, x_i, \ldots, x_N,$$

where x_i is 1 or 0 according as to whether the ith member has A or not.* Regarding the x's as quantitative variables, we have by classical results the unbiased estimates

$$\hat{p} = x. = (\Sigma x_i)/N \quad \text{and} \quad \hat{\sigma}^2 = (\Sigma(x_i - x.)^2)/(N-1).$$

If r of the x's are 1, while s are 0, we find

$$\hat{p} = r/N \quad \text{and} \quad \hat{\sigma}^2 = rs/N(N-1).$$

Using this unbiased estimate of variance in Prof. Pearson's para. 43, we get, instead of his (28·2),

$$\frac{d}{s_d} = \frac{a - rm/N}{\sqrt{\dfrac{mnrs}{N^2(N-1)}}}, \tag{1}$$

agreeing exactly with his (22).

2. To carry the argument further, in classical theory, if we have two samples

$$(x_1, x_2, \ldots, x_i, \ldots, x_m) \quad \text{and} \quad (y_1, y_2, \ldots, y_j, \ldots, y_n)$$

to test whether the samples come from the same normal population we take

$$t = \frac{x. - y.}{s}\sqrt{\frac{mn}{m+n}},$$

where $x. = (\Sigma x_i)/m$, $y. = (\Sigma y_j)/n$, and

$$s^2 = \frac{\Sigma(x_i - x.)^2 + \Sigma(y_j - y.)^2}{m+n-2}, \tag{2}$$

and use tables of the t distribution for $(m+n-2)$ degrees of freedom.

It is common practice to neglect departures from normality in applying this test. If we do so, and apply it to our qualitative case along the lines indicated above, we get

$$t = \frac{a - rm/N}{\sqrt{\dfrac{acn + bdm}{N(N-2)}}},$$

which, if we are justified in our neglect of departures from normality, should be distributed as t on $(N-2)$ degrees of freedom.

* For a similar argument see B. L. Welch (1938, p. 155).

3. To obtain the formula (1) on these lines, we have in effect to commit the well-known fallacy of replacing s^2 as given by (2), by

$$s'^2 = \frac{\Sigma(x_i - m')^2 + \Sigma(y_j - m')^2}{m+n-1}, \tag{3}$$

where $$m' = (\Sigma x_i + \Sigma y_j)/(m+n).$$

We are led to ask why (3) should be approximately correct (and in fact it is better than (2)) in the qualitative case, while (2) is preferred in the quantitative case.

4. The simplest reason for preferring (2) to (3) in the quantitative case is that s'^2 is not independent of $(x. - y.)$, so that the conditions for validity of the t distribution are not satisfied. In our qualitative case this argument loses validity, since neither s^2 nor s'^2 is independent of $(x. - y.)$.

The second reason for preferring (2) to (3) in the quantitative case is more complicated, but for our purposes it reduces essentially to the fact that, in the case of normal distributions, and *only in this case*, the mean and variance of samples are independently distributed, so that the common mean value of the populations, estimated by m', is irrelevant to the test for differences. In our qualitative case, on the other hand, m' contributes to our knowledge of the variance.

5. If we apply Pitman's 'absolute' analogue of the t test to our case, we arrive at the hypergeometric series of Prof. Pearson's Problem I. But Bartlett's argument, showing the convergence of Pitman's test and the t test, will apply here only in very large samples, because of the finite probability of obtaining observed values which coincide.

6. From the above point of view, Prof. Pearson's analysis of his Problem II may be regarded in one sense as an examination of the effect of large departures from normality on the t test. In this light, his conclusions given in paras. 51 and 52 are seen to extend to the t test, as well as to the 2×2 table problem.

7. If I may state my personal attitude, it is that statistics is a branch of applied mathematics, like symbolic logic or hydrodynamics. Examination of foundations is desirable, but it must be remembered that undue emphasis on niceties is a disease to which persons with mathematical training are specially prone. In pure mathematics itself there are disputes on foundations which closely parallel the disputes over the foundations of statistics. The lesson to be drawn is, that while statistics is a most valuable aid to judgement, it cannot wholly replace it.

8. Finally, it must be emphasized that the order of printing of Prof. Pearson's paper and my own reflects Prof. Pearson's generosity rather than the historical order of events. Much of his paper was, unknown to me, given in lectures before the war; whereas my work on the problem began only in 1943. Since then I have owed much both to Prof. Pearson's published work and to discussions which I have been privileged to have with him.

REFERENCE

Welch, B. L. (1938). *Biometrika*, 30, 155.

THE CUMULANTS OF THE Z AND OF THE LOGARITHMIC χ^2 AND t DISTRIBUTIONS

By JOHN WISHART

School of Agriculture, Cambridge

Explicit expressions for the exact cumulants of Fisher's z-distribution do not appear ever to have been published. They were therefore worked out, and appear in § 2 of this paper. It afterwards appeared that the logical method of presentation was to deal with the similar problem for $\frac{1}{2}\log(\chi^2/n)$,* since the z-distribution involves the simple difference of two such functions which are independent. This led to § 1. Since writing this paper, Bartlett & Kendall (1946) have published the same result in the form of the cumulants of $\log s^2$, and have given graphical and tabular representations for varying n up to 20. The solution is, of course, implicit in Cornish & Fisher's (1937) statement of the moment generating function, while Mr C. R. Rao has informed me that he reached the same result in work done for an M.A. Thesis of the University of Calcutta (unpublished). § 1 has accordingly been shortened, but is retained in view of the additional formulae to those of Bartlett and Kendall.

1. The logarithmic χ^2 distribution

The distribution of χ^2, for n degrees of freedom, is given by

$$\frac{1}{\Gamma(\frac{1}{2}n)}(\tfrac{1}{2}\chi^2)^{\frac{1}{2}n-1}e^{-\frac{1}{2}\chi^2}d(\tfrac{1}{2}\chi^2).$$

As pointed out by Cornish & Fisher (1937), the mean value of $\exp\{\frac{1}{2}it\log(\chi^2/n)\}$

$$\text{i.e. of } \exp\{\tfrac{1}{2}it\log(\tfrac{1}{2}\chi^2)-\tfrac{1}{2}it\log(\tfrac{1}{2}n)\}$$

is the moment generating function of the distribution of $\frac{1}{2}\log(\chi^2/n)$, namely

$$M=\frac{\Gamma\frac{1}{2}(n+it)}{\Gamma(\frac{1}{2}n)}\exp\{-\tfrac{1}{2}it\log(\tfrac{1}{2}n)\}.$$

The cumulant generating function is

$$K=\log M=-\tfrac{1}{2}it\log(\tfrac{1}{2}n)+\log\Gamma\tfrac{1}{2}(n+it)-\log\Gamma(\tfrac{1}{2}n).$$

The cumulants of the distribution of $\frac{1}{2}\log(\chi^2/n)$ are readily written down by differentiating K successively with respect to it and at each stage putting $t=0$. We have in fact

$$\begin{aligned}\kappa_1&=-\tfrac{1}{2}\log(\tfrac{1}{2}n)+\frac{d}{dn}\log\Gamma(\tfrac{1}{2}n)\\&=-\frac{1}{2}\left\{\log a+\mathrm{Lt}_{s\to1}\left(\zeta(s,a)-\frac{1}{s-1}\right)\right\}\end{aligned}\qquad(1)$$

and

$$\kappa_s=\frac{(-1)^s(s-1)!}{2^s}\zeta(s,a)\quad(s>1,\ a=\tfrac{1}{2}n),$$

where $\zeta(s,a)$ denotes the generalized Zeta-function

$$\zeta(s,a)=\sum_{j=0}^{\infty}\frac{1}{(a+j)^s}.$$

* All logarithms in this paper are to base e.

The cumulants may be readily computed by throwing them into the form

$$2\kappa_1 = \psi(\tfrac{1}{2}n) - \log(\tfrac{1}{2}n),$$

$$2^s\kappa_s = \psi^{(s-1)}(\tfrac{1}{2}n), \tag{2}$$

where $\psi(x) = d\{\log \Gamma(x)\}/dx, \quad \psi^{(s-1)}(x) = d^s\{\log \Gamma(x)\}/dx^s.$

$\psi(x)$ is variously called the Psi or Digamma function, and its derivatives have been called the Trigamma, Tetragamma, etc. Functions, and the series the Polygamma Functions. These functions have been computed in some considerable detail. For n up to 22 the mean and variance can be got from Elinor Pairman's 'Tables of the Digamma and Trigamma functions' (1919). Tables up to Pentagamma appear in Vol. I of the British Association's *Mathematical Tables* (1931), but with certain gaps which, although intended to be bridged by reduction formulae, render the tables less generally useful (for n less than 22) than H. T. Davis's Tables (1933, 1935). Table 10 of Vol. I gives all that is required for $\psi(x)$; in Vol. II, Tables 14–16, 18–20, 22–24 and 26–28 cover a wide range up to Hexagamma.

As shown by Bartlett & Kendall (1946), the approach to normality is very slow. For $n = 24$ (the limit for n_1 of the z table of Fisher & Yates (1943), which provides percentage points for the distribution under consideration in the line $n_2 = \infty$) the cumulants have been worked out to κ_6, the last being specially computed from its formula given below. The gamma ratios are $\gamma_1 = -0{\cdot}295$, $\gamma_2 = 0{\cdot}174$, $\gamma_3 = -0{\cdot}154$ and $\gamma_4 = 0{\cdot}175$, and $|\gamma|$ increases thereafter at this level of n instead of tending to zero. Approximate percentage points may, however, be worked out by using the formulae at the foot of the z table, putting $n_2 = \infty$.

For small n, we note that

$$\zeta(s) = \sum_{j=1}^{\infty} \frac{1}{j^s} = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \dots + \frac{1}{(r-1)^s} + \zeta(s, r) \quad r \text{ an integer},$$

$$\zeta(s).(1-2^{-s}) = \frac{1}{1^s} + \frac{1}{3^s} + \frac{1}{5^s} + \dots + \infty$$

$$= \frac{1}{1^s} + \frac{1}{3^s} + \frac{1}{5^s} + \dots + \frac{1}{(2r-1)^s} + 2^{-s}\zeta(s, r+\tfrac{1}{2}).$$

We thus get, *for* $n = 2r$,

$$\kappa_s = \frac{(-1)^s(s-1)!}{2^s}\left\{\zeta(s) - \left(1 + \frac{1}{2^s} + \frac{1}{3^s} + \dots + \frac{1}{(\frac{1}{2}n-1)^s}\right)\right\} \quad (s>1), \tag{3}$$

in which the terms in $\{\dots\}$ reduce to $\zeta(s)$ for $n = 2$.

For $n = 2r+1$

$$\kappa_s = (-1)^s(s-1)!\left\{\zeta(s)(1-2^{-s}) - \left(1 + \frac{1}{3^s} + \frac{1}{5^s} + \dots + \frac{1}{(n-2)^s}\right)\right\} \quad (s>1), \tag{4}$$

in which the terms in $\{\dots\}$ reduce to $\zeta(s)(1-2^{-s})$ for $n = 1$.

In the special case of $s = 1$ we have

$$n = 2r \qquad \kappa_1 = -\tfrac{1}{2}(\gamma + \log\tfrac{1}{2}n) + \frac{1}{2}\left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{\frac{1}{2}n-1}\right),$$

$$n = 2r+1 \qquad \kappa_1 = -\tfrac{1}{2}(\gamma + \log 2n) + \left(1 + \frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{n-2}\right). \tag{5}$$

For $n = 2$ and 1 respectively these expressions reduce to the first bracket. $\zeta(s)$ can be got from tables, and in particular

$$\zeta(2m) = 2^{2m-1}\pi^{2m}B_m/(2m)!,$$

where the B's are the Bernoulli numbers $B_1 = \frac{1}{6}$, $B_2 = \frac{1}{30}$, $B_3 = \frac{1}{42}$, $B_4 = \frac{1}{30}$, $B_5 = \frac{5}{66}$, etc. γ is Euler's constant. For reference we may quote:

$$\gamma = 0{\cdot}57721\,56649, \qquad \zeta(4) = 1{\cdot}08232\,32337,$$
$$\zeta(2) = 1{\cdot}64493\,40668, \qquad \zeta(5) = 1{\cdot}03692\,77551,$$
$$\zeta(3) = 1{\cdot}20205\,69032, \qquad \zeta(6) = 1{\cdot}01734\,30620.$$

Note that $$\mathrm{Lt}_{s\to 1}\left(\zeta(s,a)-\frac{1}{s-1}\right) = \gamma+\sum_{j=0}^{\infty}\left(\frac{1}{a+j}-\frac{1}{1+j}\right), \quad (R(a)>0).$$

For large n, asymptotic formulae for the Zeta-function may be used, and we get

$$\kappa_1 \sim -\frac{1}{2n}-\sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j}{jn^{2j}}$$
$$\sim -\frac{1}{2n}-\frac{1}{6n^2}+\frac{1}{15n^4}-\frac{8}{63n^6}+\frac{8}{15n^8}-\frac{64}{33n^{10}}+\dots, \tag{6}$$
$$\kappa_s \sim (-1)^s\left[\frac{(s-2)!}{2n^{s-1}}+\frac{(s-1)!}{2n^s}+\frac{2}{n^s}\sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j(2j+s-2)!}{(2j)!\,n^{2j-1}}\right] \quad (s>1).$$

We may note in passing that not only may this general expression for κ_s be applied to the special case $s = 1$ with the proviso that the first term in that case is dropped, but also that κ_2 may be obtained from $\kappa_1+\frac{1}{2}\log(\frac{1}{2}n)$ by term-by-term differentiation with respect to n, and likewise κ_3 from κ_2, κ_4 from κ_3, etc., by similar term-by-term differentiation. This follows from a property of the Zeta-function. It is therefore not necessary to write down the explicit expressions for κ_2, κ_3, etc., but we may note that their leading terms are $\frac{1}{2}n^{-1}$, $-\frac{1}{2}n^{-2}$, n^{-3}, $-3n^{-4}$, $12n^{-5}$, etc., so that the leading terms of γ_1 and γ_2 are $-\sqrt{(2/n)}$ and $4/n$ respectively, while γ_r is $O(n^{-\frac{1}{2}r})$. More exactly we have, writing $n' = n-1$,

$$\kappa_2 = \frac{1}{2n'}\left(1-\frac{1}{3n'^2}+\frac{7}{15n'^4}+O\left(\frac{1}{n'^6}\right)\right)$$

with corresponding expressions for κ_3, κ_4, etc., obtained by differentiation with respect to n', and

$$\gamma_1 = -\sqrt{\frac{2}{n'}}\left(1-\frac{1}{2n'^2}+O\left(\frac{1}{n'^4}\right)\right), \qquad \gamma_r \sim (-1)^r r!\left(\frac{2}{n'}\right)^{\frac{1}{2}r},$$
$$\gamma_2 = \frac{4}{n'}\left(1-\frac{4}{3n'^2}+O\left(\frac{1}{n'^4}\right)\right), \qquad \frac{\gamma_r}{\gamma_{r-1}} \sim -r\sqrt{\left(\frac{2}{n'}\right)}.$$

Finally, if instead of the distribution of $\frac{1}{2}\log(\chi^2/n)$ we are interested in the distribution of $\log(s^2)$, where s^2 is an estimate of σ^2 based on n degrees of freedom, we have

$$\log(\chi^2/n) = \log s^2 - \log\sigma^2,$$

and thus for the distribution of $\log(s^2)$ we have

$$\kappa_1 = \log\left(\frac{2\sigma^2}{n}\right)+2\frac{d}{dn}\log\Gamma(\tfrac{1}{2}n)$$
$$= \log\left(\frac{2\sigma^2}{n}\right)-\mathrm{Lt}_{s\to 1}\left\{\zeta(s,a)-\frac{1}{s-1}\right\}$$

and $$\kappa_s = (-1)^s(s-1)!\,\zeta(s,a) \quad (s>1,\ a=\tfrac{1}{2}n),$$

while the γ ratios are the same as for $\frac{1}{2}\log(\chi^2/n)$. Obviously $\log\chi$ and $\log s$ can be treated similarly. See Bartlett & Kendall (1946).

2. THE z DISTRIBUTION

The distribution of $z = \frac{1}{2}\log(s_1^2/s_2^2)$, where s_1^2 and s_2^2 are independent estimates of a variance σ^2, based respectively on ν_1 and ν_2 degrees of freedom, is obviously that of

$$\tfrac{1}{2}\log(\chi_1^2/\nu_1) - \tfrac{1}{2}\log(\chi_2^2/\nu_2)$$

and its cumulants may therefore be at once derived from those of the logarithmic χ^2 distribution. The cumulant generating function is

$$K = \log M = \tfrac{1}{2}it\log(\nu_2/\nu_1) + \log\Gamma\tfrac{1}{2}(\nu_1+it) + \log\Gamma\tfrac{1}{2}(\nu_2-it) - \log\Gamma(\tfrac{1}{2}\nu_1) - \log\Gamma(\tfrac{1}{2}\nu_2).$$

Further, we have

$$\begin{aligned}\kappa_1 &= \tfrac{1}{2}\log\frac{\nu_2}{\nu_1} + \frac{d}{d\nu_1}\log\Gamma(\tfrac{1}{2}\nu_1) - \frac{d}{d\nu_2}\log\Gamma(\tfrac{1}{2}\nu_2)\\ &= \frac{1}{2}\left\{\log\frac{\nu_2}{\nu_1} + \mathrm{Lt}_{s\to 1}(\zeta(s,a_2)-\zeta(s,a_1))\right\} \qquad (7)\\ \kappa_s &= 2^{-s}(s-1)!\,\{\zeta(s,a_2)+(-1)^s\zeta(s,a_1)\} \quad (s>1,\ a_1=\tfrac{1}{2}\nu_1,\ a_2=\tfrac{1}{2}\nu_2).\end{aligned}$$

For computing purposes these may be thrown into the forms

$$\begin{aligned}2\kappa_1 &= \log(\nu_2/\nu_1) + \psi(\tfrac{1}{2}\nu_1) - \psi(\tfrac{1}{2}\nu_2),\\ 2^s\kappa_s &= \psi^{(s-1)}(\tfrac{1}{2}\nu_1) + (-1)^s\psi^{(s-1)}(\tfrac{1}{2}\nu_2) \quad (s>1). \qquad (8)\end{aligned}$$

To illustrate, let us take $\nu_1 = 24$, $\nu_2 = 60$. We then have from the Polygamma tables (except for κ_6, which was specially computed):

$$\begin{aligned}&\kappa_1 = -0{\cdot}0127\,429, \quad \kappa_3 = -0{\cdot}0007\,998, \quad \kappa_5 = -0{\cdot}0000\,104,\\ &\kappa_2 = 0{\cdot}0301\,992, \quad \kappa_4 = 0{\cdot}0000\,867, \quad \kappa_6 = 0{\cdot}0000\,019,\\ &\qquad\qquad \sigma = \sqrt{\kappa_2} = 0{\cdot}1737\,792,\end{aligned}$$

$\gamma_1 = -0{\cdot}152$, $\gamma_2 = 0{\cdot}095$ (or $\beta_1 = 0{\cdot}023$, $\beta_2 = 3{\cdot}095$), indicating the degree and nature of the departure from normality. γ_3 and γ_4 are $-0{\cdot}066$ and $0{\cdot}067$ respectively.

If as a first approximation we assume that for ν_1 and ν_2 of the order of the numbers chosen in this example, or higher, z is distributed normally with mean and variance given by the above formulae, we obtain approximate percentage points, e.g. for the 95 and 5 % points we can subtract and add $1{\cdot}6449\sigma$ from and to the mean. The result in the present case is to give us $-0{\cdot}299$ and $0{\cdot}273$, the correct values being $-0{\cdot}306$ and $0{\cdot}265$. The approximation is adequate to almost two figure accuracy, and is evidently useful when we only require to know whether an observed z is significant or not. A better approximation is provided by the formulae attached to the z tables (see Fisher & Yates (1943)), which yield $-0{\cdot}3045$ and $0{\cdot}2653$ as against the correct values of $-0{\cdot}3055$ (see Thompson (1941)) and $0{\cdot}2654$.

Explicit algebraic expressions are readily written down for the cumulants for small ν_1 and ν_2, using the same method as for the logarithmic χ^2 distribution. Where it is necessary to do so, ν_1 will be assumed less than ν_2. In the contrary case we need only interchange ν_1 and ν_2, changing the sign of the odd cumulants in so doing. The odd cumulants are zero when $\nu_1 = \nu_2$. We have

Even cumulants $\qquad (r = 2s,\ s > 0)$

$\nu_1 = 2p$, $\nu_2 = 2q$

$$\kappa_r = 2(r-1)!\left[\zeta(r)\,2^{-r} - \left(\frac{1}{2^r}+\frac{1}{4^r}+\dots+\frac{1}{(\nu_1-2)^r}\right) - \frac{1}{2}\left(\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-2)^r}\right)\right], \qquad (9)$$

$\nu_1 = 2p+1$, $\nu_2 = 2q+1$

$$\kappa_r = 2(r-1)!\left[\zeta(r)(1-2^{-r}) - \left(\frac{1}{1^r}+\frac{1}{3^r}+\dots+\frac{1}{(\nu_1-2)^r}\right) - \frac{1}{2}\left(\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-2)^r}\right)\right]. \quad (10)$$

$\nu_1 = \nu_2$. Drop out the last bracket of terms in the above two cases.

$\nu_1 = 2p$, $\nu_2 = 2q+1$

$$\kappa_r = (r-1)!\left[\zeta(r) - \left(\frac{1}{2^r}+\frac{1}{4^r}+\dots+\frac{1}{(\nu_1-2)^r}\right) - \left(1+\frac{1}{3^r}+\frac{1}{5^r}+\dots+\frac{1}{(\nu_2-2)^r}\right)\right]. \quad (11)$$

$\nu_1 = 2p+1$, $\nu_2 = 2q$. Interchange ν_1 and ν_2 in this last case.

Odd cumulants $(r = 2s+1,\ s>0)$

$\nu_1 = 2p$, $\nu_2 = 2q$, or $\nu_1 = 2p+1$, $\nu_2 = 2q+1$

$$\kappa_r = -(r-1)!\left[\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-2)^r}\right]. \quad (12)$$

$\nu_1 = 2p$, $\nu_2 = 2q+1$

$$\kappa_r = (r-1)!\left[\zeta(r)(1-2^{1-r}) + \left(\frac{1}{2^r}+\frac{1}{4^r}+\dots+\frac{1}{(\nu_1-2)^r}\right) - \left(1+\frac{1}{3^r}+\frac{1}{5^r}+\dots+\frac{1}{(\nu_2-2)^r}\right)\right]. \quad (13)$$

$\nu_1 = 2p+1$, $\nu_2 = 2q$. Interchange ν_1 and ν_2 in this last case, and change the sign of κ_r.

In the special case of $s = 1$, we have

$\nu_1 = 2p$, $\nu_2 = 2q$, or $\nu_1 = 2p+1$, $\nu_2 = 2q+1$

$$\kappa_1 = \tfrac{1}{2}\log\left(\frac{\nu_2}{\nu_1}\right) - \left(\frac{1}{\nu_1}+\frac{1}{\nu_1+2}+\dots+\frac{1}{\nu_2-2}\right). \quad (14)$$

$\nu_1 = 2p$, $\nu_2 = 2q+1$

$$\kappa_1 = \tfrac{1}{2}\log\left(\frac{\nu_2}{\nu_1}\right) + \log 2 + \left(\frac{1}{2}+\frac{1}{4}+\dots+\frac{1}{\nu_1-2}\right) - \left(1+\frac{1}{3}+\frac{1}{5}+\dots+\frac{1}{\nu_2-2}\right). \quad (15)$$

$\nu_1 = 2p+1$, $\nu_2 = 2q$. Interchange ν_1 and ν_2 in this last case and change the sign of κ_1.

For large ν_1 *and* ν_2, a combination of the asymptotic formulae already given readily yields the following results:

$$\kappa_1 \sim \frac{1}{2}\left(\frac{1}{\nu_2}-\frac{1}{\nu_1}\right) + \sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j}{j}\left(\frac{1}{\nu_2^{2j}}-\frac{1}{\nu_1^{2j}}\right),$$

the numerical coefficients being as for the κ_1 of $\tfrac{1}{2}\log(\chi^2/n)$,

$$\kappa_s \sim \frac{(s-2)!}{2}\left(\frac{1}{\nu_2^{s-1}}+\frac{(-1)^s}{\nu_1^{s-1}}\right) + \frac{(s-1)!}{2}\left(\frac{1}{\nu_2^s}+\frac{(-1)^s}{\nu_1^s}\right)$$
$$+2\sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j(2j+s-2)!}{(2j)!}\left(\frac{1}{\nu_2^{2j+s-1}}+\frac{(-1)^s}{\nu_1^{2j+s-1}}\right). \quad (16)$$

We may put $s = 1$ in κ_s provided we drop the first term. We note also that κ_2 and higher cumulants can be written down immediately by differentiating the terms in ν_2 and ν_1 of $\kappa_1 - \tfrac{1}{2}\log(\nu_2/\nu_1)$ successively with respect to $-\nu_2$ and ν_1 respectively.

These are the results given by Cornish & Fisher (1937), whose formulae can be extended at sight by means of the results of this paper. A first approximation not only gives the familiar results

$$\kappa_1 \sim \frac{1}{2}\left(\frac{1}{\nu_2}-\frac{1}{\nu_1}\right),\quad \kappa_2 \sim \frac{1}{2}\left(\frac{1}{\nu_2}+\frac{1}{\nu_1}\right),$$

but also the more general $\quad \kappa_s \sim \dfrac{(s-2)!}{2}\left(\dfrac{1}{\nu_2^{s-1}}+\dfrac{(-1)^s}{\nu_1^{s-1}}\right),$

but it should be noted that for all $s > 1$ a second approximation, which takes in an additional term, is

$$\kappa_s \sim \frac{(s-2)!}{2}\left(\frac{1}{(\nu_2-1)^{s-1}}+\frac{(-1)^s}{(\nu_1-1)^{s-1}}\right).$$

The accuracy of the asymptotic approximation at the limits of the z table given by Fisher & Yates (1943) can be seen by applying it to our example ($\nu_1 = 24$, $\nu_2 = 60$). The numbers of terms which are significant in the eighth place (needed for final accuracy to 7 decimal places), are three for κ_1, four for κ_2 and κ_3, and three for κ_4, κ_5 and κ_6. The first term for κ_6, namely $12(\nu_2^{-5}+\nu_1^{-5})$, yields 0·0000 015, rather more than 20 % too low. To use

$$12\{(\nu_2-1)^{-5}+(\nu_1-1)^{-5}\}$$

would give 0·0000 019, about 2 % too high.

Should ν_1 or ν_2 be only moderate in size, the other being large, we may make use of the relation

$$\zeta(s,a) = \frac{1}{a^s}+\frac{1}{(a+1)^s}+\ldots+\frac{1}{(a+r-1)^s}+\zeta(s,a+r), \quad (r \text{ an integer}),$$

where a is one-half of the smaller of ν_1 or ν_2, to convert our formulae into forms in which asymptotic expansions may be applied to both of the Zeta-functions. We then have ($\nu_1 < \nu_2$):

$$\kappa_1 = \frac{1}{2}\left[\log\left(\frac{\nu_2}{\nu_1}\right)+\mathrm{Lt}_{s\to 1}\{\zeta(s,\tfrac{1}{2}\nu_2)-\zeta(s,\tfrac{1}{2}\nu_1+r)\}\right]-\left(\frac{1}{\nu_1}+\frac{1}{\nu_1+2}+\ldots+\frac{1}{\nu_1+2r-2}\right),$$

$$\kappa_s = (s-1)!\left[2^{-s}\{\zeta(s,\tfrac{1}{2}\nu_2)+(-1)^s\zeta(s,\tfrac{1}{2}\nu_1+r)\}+(-1)^s\left\{\frac{1}{\nu_1^s}+\frac{1}{(\nu_1+2)^s}+\ldots+\frac{1}{(\nu_1+2r-2)^s}\right\}\right], \tag{17}$$

and

$$\zeta(s,n) \sim \frac{1}{(s-1)\,n^{s-1}}+\frac{1}{2n^s}+\frac{1}{(s-1)!\,n^s}\sum_{j=1}^{\infty}\frac{(-1)^{j-1}B_j(2j+s-2)!}{(2j)!\,n^{2j-1}}.$$

Particular cases of some interest arise (i) when $r = \frac{1}{2}(\nu_2-\nu_1)$, ν_1 and ν_2 being either both odd or both even, and (ii) when $r = \frac{1}{2}(\nu_2-\nu_1+1)$, ν_1 (or ν_2) being even and ν_2 (or ν_1) odd. In the former case the first term within squared brackets in κ_s is $2^{-s}(1+(-1)^s)\,\zeta(s,\frac{1}{2}\nu_2)$, which is zero when s is odd and $2^{1-s}\zeta(s,\frac{1}{2}\nu_2)$ when s is even. In the latter we have

$$2^{-s}\{\zeta(s,\tfrac{1}{2}\nu_2)+(-1)^s\zeta(s,\tfrac{1}{2}(\nu_2+1))\}$$

which is $\zeta(s,\nu_2)$ when s is even. With s odd we are concerned with the difference of two Zeta-functions in which the a's differ by one-half, and the expression may be written

$$\sum_{j=0}^{\infty}\frac{(-1)^j}{(\nu_2+j)^s} = \frac{1}{(s-1)!}\left(\frac{d}{d\nu_2}\right)^{s-1}\sum_{j=0}^{\infty}\frac{(-1)^j}{\nu_2+j}$$

and

$$\begin{aligned}\sum_{j=0}^{\infty}\frac{(-1)^j}{\nu_2+j} &= \int_0^1\frac{x^{\nu_2-1}\,dx}{1+x}\\ &= \sum_{j=0}^{\infty}\frac{j!(\nu_2-1)!}{2^{j+1}(\nu_2+j)!} \quad \text{on integration by parts}\\ &= \frac{1}{2\nu_2}+\sum_{j=1}^{\infty}\frac{(-1)^{j-1}(2^{2j}-1)B_j}{2j\,\nu_2^{2j}}\end{aligned}$$

on expansion in powers of ν_2^{-1}. This asymptotic expansion is an interesting one in which the early coefficients are very simple, for the series is

$$\frac{1}{2\nu_2}+\frac{1}{4\nu_2^2}-\frac{1}{8\nu_2^4}+\frac{1}{4\nu_2^6}-\frac{17}{16\nu_2^8}+\ldots. \tag{18}$$

The various cases are set out below:

Even cumulants $\quad (r = 2s,\ s > 0)$

$\nu_1 = 2p,\ \nu_2 = 2q$, or $\nu_1 = 2p+1,\ \nu_2 = 2q+1$

$$\kappa_r = (r-1)!\left\{\frac{1}{\nu_1^r}+\frac{1}{(\nu+2)^r}+\dots+\frac{1}{(\nu_2-2)^r}\right\} + \frac{(r-2)!}{\nu_2^{r-1}}+\frac{(r-1)!}{\nu_2^r}-\frac{1}{\nu_2^r}\sum_{j=1}^{\infty}\frac{(-4)^j B_j(2j+r-2)!}{(2j)!\,\nu_2^{2j-1}}, \tag{19}$$

$\nu_1 = 2p,\ \nu_2 = 2q+1$

$$\kappa_r = (r-1)!\left\{\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-1)^r}\right\} + \frac{(r-2)!}{\nu_2^{r-1}}+\frac{(r-1)!}{2\nu_2^r}-\frac{1}{\nu_2^r}\sum_{j=1}^{\infty}\frac{(-1)^j B_j(2j+r-2)!}{(2j)!\,\nu_2^{2j-1}}. \tag{20}$$

Odd cumulants $\quad (r = 2s+1,\ s > 0)$

$\nu_1 = 2p,\ \nu_2 = 2q$, or $\nu_1 = 2p+1,\ \nu_2 = 2q+1$

$$\kappa_r = -(r-1)!\left\{\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-2)^r}\right\}, \tag{21}$$

$\nu_1 = 2p,\ \nu_2 = 2q+1$

$$\kappa_r = -(r-1)!\left\{\frac{1}{\nu_1^r}+\frac{1}{(\nu_1+2)^r}+\dots+\frac{1}{(\nu_2-1)^r}\right\} + \frac{(r-1)!}{2\nu_2^r}+\frac{1}{\nu_2^r}\sum_{j=1}^{\infty}\frac{(-1)^{j-1}(2^{2j}-1)B_j(2j+r-2)!}{(2j)!\,\nu_2^{2j-1}}. \tag{22}$$

In the special case of $s = 1$, we have

$\nu_1 = 2p,\ \nu_2 = 2q$, or $\nu_1 = 2p+1,\ \nu_2 = 2q+1$

$$\kappa_1 = \tfrac{1}{2}\log\left(\frac{\nu_2}{\nu_1}\right)-\left(\frac{1}{\nu_1}+\frac{1}{\nu_1+2}+\dots+\frac{1}{\nu_2-2}\right),$$

$\nu_1 = 2p,\ \nu_2 = 2q+1$

$$\kappa_1 = \tfrac{1}{2}\log\left(\frac{\nu_2}{\nu_1}\right)-\left(\frac{1}{\nu_1}+\frac{1}{\nu_1+2}+\dots+\frac{1}{\nu_2-1}\right)+\frac{1}{2\nu_2}+\sum_{j=1}^{\infty}\frac{(-1)^{j-1}(2^{2j}-1)B_j}{2j\nu_2^{2j}}. \tag{23}$$

3. The logarithmic t-distribution

When $\nu_1 = 1$. $z = \log|t|$, and we thus have as a special case for the distribution of $\log|t|$ for $\nu_2 = n$ degrees of freedom:

$$2\kappa_1 = \log n + \mathrm{Lt}_{s\to 1}\{\zeta(s,\tfrac{1}{2}n)-\zeta(s,\tfrac{1}{2})\} \tag{24}$$

$$= \log n + \psi(\tfrac{1}{2})-\psi(\tfrac{1}{2}n) \qquad (\psi(\tfrac{1}{2}) = -\gamma-2\log 2). \tag{25}$$

For small n

$n = 2p$
$$\kappa_1 = \tfrac{1}{2}\log n - \log 2 - \left(\frac{1}{2}+\frac{1}{4}+\dots+\frac{1}{n-2}\right),$$

$n = 2p+1$
$$\kappa_1 = \tfrac{1}{2}\log n - \left(1+\frac{1}{3}+\frac{1}{5}+\dots+\frac{1}{n-2}\right). \tag{26}$$

For large n

$$\kappa_1 \sim -\tfrac{1}{2}(\gamma+\log 2)+\frac{1}{2n}+\sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j}{jn^{2j}}. \tag{27}$$

Also $$2^s\kappa_s = (s-1)!\{\zeta(s,\tfrac{1}{2}n)+(-1)^s\zeta(s,\tfrac{1}{2})\} \quad (s>1), \tag{28}$$
$$= \psi^{(s-1)}(\tfrac{1}{2})+(-1)^s\psi^{(s-1)}(\tfrac{1}{2}n) \qquad (\psi^{(s-1)}(\tfrac{1}{2}) = (-1)^s(s-1)!(2^s-1)\zeta(s)). \tag{29}$$

For small n we have the following cases:

Even cumulants $(r = 2s,\ s>0)$

$$n = 2p \qquad \kappa_r = (r-1)!\left\{\zeta(r)-\left(\frac{1}{2^r}+\frac{1}{4^r}+\dots+\frac{1}{(n-2)^r}\right)\right\}, \tag{30}$$

$$n = 2p+1 \qquad \kappa_r = (r-1)!\left\{2\zeta(r)(1-2^{-r})-\left(1+\frac{1}{3^r}+\frac{1}{5^r}+\dots+\frac{1}{(n-2)^r}\right)\right\}.$$

Odd cumulants $(r = 2s+1,\ s>0)$

$$n = 2p \qquad \kappa_r = -(r-1)!\left\{\zeta(r)(1-2^{1-r})+\left(\frac{1}{2^r}+\frac{1}{4^r}+\dots+\frac{1}{(n-2)^r}\right)\right\}, \tag{31}$$

$$n = 2p+1 \qquad \kappa_r = -(r-1)!\left(1+\frac{1}{3^r}+\frac{1}{5^r}+\dots+\frac{1}{(n-2)^r}\right).$$

For large n

$$\kappa_s \sim (-1)^s(s-1)!\zeta(s)(1-2^{-s})+\frac{(s-2)!}{2n^{s-1}}+\frac{(s-1)!}{2n^s}+\frac{2}{n^s}\sum_{j=1}^{\infty}\frac{(-4)^{j-1}B_j(2j+s-2)!}{(2j)!\,n^{2j-1}} \quad (s>1). \tag{32}$$

In the special case of $n=\infty$ we have for the distribution of $\log|x|$, where x is a normal variable with zero mean and unit standard deviation:

$$\kappa_1 = -\tfrac{1}{2}(\gamma+\log 2), \quad \kappa_s = (-1)^s(s-1)!\zeta(s)(1-2^{-s}), \tag{33}$$

as follows also from the case of $\frac{1}{2}\log(\chi^2/n)$ on putting $n=1$.

4. NOTE ON THE χ^2 DISTRIBUTION APPROXIMATION

Fisher's result that $\sqrt{(2\chi^2)}$ is approximately normally distributed about a mean of $\sqrt{(2n-1)}$ with unit variance (n being the number of degrees of freedom) is well known. The demonstration depends on showing that the mean value of χ is

$$\kappa_1 = \sqrt{2}\,\Gamma\tfrac{1}{2}(n+1)/\Gamma(\tfrac{1}{2}n) \sim \sqrt{(n-\tfrac{1}{2})} \quad \text{for large } n$$

and that the variance is $$n-\kappa_1^2 \sim \tfrac{1}{2},$$

but to this order of approximation it is not possible to show that γ_1 and γ_2 tend to zero with increasing n. A formula for the ratio of the two Gamma functions, developed as far as terms in n^{-3} (see Wishart (1925)), gives $\gamma_1 \sim (2n)^{-\frac{1}{2}}$ and $\gamma_2 = O(n^{-2})$ (see, for example, Kendall (1945)), but owing to the vanishing of the term in n^{-1} of γ_2 its leading term has so far not been accurately obtained, although the exact (but somewhat complicated) expressions for the β_1 and β_2 of the distribution of $s = \sigma\chi/\sqrt{(n+1)}$ were given in an editorial in *Biometrika* (1915), **10**, 522.

Since $$\tfrac{1}{2}\{\psi\tfrac{1}{2}(n+1)-\psi(\tfrac{1}{2}n)\} = \frac{1}{2n}+\sum_{j=1}^{\infty}\frac{(-1)^{j-1}(2^{2j}-1)B_j}{2jn^{2j}},$$

by the formula given in § 2, we find on integration, and insertion of the appropriate constant, that

$$\log\Gamma\tfrac{1}{2}(n+1)-\log\Gamma(\tfrac{1}{2}n) = \tfrac{1}{2}\log(\tfrac{1}{2}n)+\sum_{j=1}^{\infty}\frac{(-1)^j(2^{2j}-1)B_j}{2j(2j-1)\,n^{2j-1}},$$

and thus have $$\frac{\Gamma\frac{1}{2}(n+1)}{\Gamma(\frac{1}{2}n)} = \sqrt{(\tfrac{1}{2}n)}\exp-\frac{1}{4n}\left\{1+\sum_{j=2}^{\infty}\frac{(-1)^{j-1}(2^{2j}-1)B_j}{j(j-\frac{1}{2})\,n^{2(j-1)}}\right\}, \tag{34}$$

which can readily be expanded to give the additional terms necessary to enable the cumulants of χ (or of $\sqrt{(2\chi^2)}$) to be worked out (see Johnson & Welch (1939)). Taking $\sqrt{\{\frac{1}{2}(n-\frac{1}{2})\}}$ as the first approximation we find

$$\begin{aligned}\frac{\Gamma\frac{1}{2}(n+1)}{\Gamma(\frac{1}{2}n)} &= \sqrt{\left(\frac{n-\frac{1}{2}}{2}\right)} \exp\left\{\frac{1}{16n^2}\left(1+\frac{1}{n}+\frac{1}{8n^2}-\frac{3}{4n^3}+\ldots\right)\right\} \\ &= \sqrt{\left(\frac{n-\frac{1}{2}}{2}\right)}\left(1+\frac{1}{16n(n-1)}\right)+O(n^{-3\cdot5}),\end{aligned} \tag{35}$$

thus providing a second approximation to the ratio of two Gamma functions differing by one-half. The cumulants of $\sqrt{(2\chi^2)}$ are

$$\begin{aligned}\kappa_1 &= \sqrt{(2n-1)}\left(1+\frac{1}{16n(n-1)}\right)+O(n^{-3\cdot5}), \\ \kappa_2 &= 1-\frac{1}{4n}-\frac{1}{8n^2}+\frac{5}{64n^3}-O(n^{-4}), \\ \kappa_3 &= \frac{1}{\sqrt{(2n)}}\left(1+\frac{1}{4n}-\frac{13}{32n^2}\right)+O(n^{-3\cdot5}), \\ \kappa_4 &= \frac{3}{4n^2}\left(1+\frac{1}{n}\right)+O(n^{-4}),\end{aligned} \tag{36}$$

so that

$$\begin{aligned}\gamma_1 &= \frac{1}{\sqrt{(2n)}}\left(1+\frac{5}{8n}-\frac{1}{128n^2}\right)+O(n^{-3\cdot5}), \\ \gamma_2 &= \frac{3}{4n^2}\left(1+\frac{3}{2n}\right)+O(n^{-4}).\end{aligned}$$

The Editorial in *Biometrika* (1915), **10**, 523 calls attention in a footnote to 'Student's' approximations for the β_1 and β_2 of the sample standard deviation. The above formulae show that 'Student's' results should be

$$\begin{aligned}\beta_1 &= \frac{1}{2n}\left(1+\frac{9}{4n}+\frac{31}{8n^2}\right)+O(n^{-4}), \\ \beta_2 &= 3\left(1+\frac{1}{4n^2}+\frac{7}{8n^3}\right)+O(n^{-4}),\end{aligned}$$

in which n is now the size of the sample. For $n = 10$ these give values too low by 2 and 5 respectively in the fourth place of decimals. Practically four-figure accuracy can be attained with n as low as 10 if in the terms in n^{-3} we replace 31/8 by 17/4 and 7/8 by 1.

REFERENCES

BARTLETT, M. S. & KENDALL, D. G. (1946). *J.R. Statist. Soc. Suppl.* **8**, 128.

BRITISH ASSOCIATION (1931). *Mathematical Tables*, **1**, 42. London: Cambridge University Press.

CORNISH, E. A. & FISHER, R. A. (1937). *Rev. Inst. Internat. Statist.* **4**, 1.

DAVIS, H. T. (1933, 1935). *Tables of the Higher Mathematical Functions*, **1**, **2**. Indiana: Principia Press.

FISHER, R. A. & YATES, F. (1943). *Statistical Tables*, 2nd ed. London: Oliver and Boyd.

JOHNSON, N. L. & WELCH, B. L. (1939). *Biometrika*, **31**, 216.

KENDALL, M. G. (1945). *Advanced Theory of Statistics*, **1**, 2nd ed., § 12·7. London: Griffin and Co.

PAIRMAN, E. (1919). *Tracts for Computers*, no. 1. London: Cambridge University Press.

THOMPSON, C. M. (1941). *Biometrika*, **32**, 168.

WISHART, J. (1925). *Biometrika*, **17**, 68.

THE MEANING OF A SIGNIFICANCE LEVEL

By G. A. BARNARD

A level of significance is a probability. To say that a given result is significant on the 5 % level means that some class of events has probability 0·05. Now whatever theory we may hold as to the nature of probability, in order to give a statement of probability a precise meaning we must refer to some reference class, or set of data, on which the probability is calculated. What is the reference class involved in a level of significance?

To many people the answer to this question seems simple enough. The reference class involved is the set of indefinite (possibly imaginary) repetitions of the experiment which gave the result in question. Otherwise put, the data, on which the probability is calculated, are the external conditions of the experiment. The following example indicates, however, that the meaning of this reference class is not always clear. The example is a modified form of one given by Prof. R. A. Fisher in a letter to the author.

Suppose we have a bag of chrysanthemum seeds, known to give plants having white flowers or plants having purple flowers, no other colours being possible. We suspect that the proportions of white and purple seeds are equal, and to test this hypothesis we select at random ten seeds from the bag, and plant them. Nine of the plants grow to maturity, and all of them have white flowers. On what level of significance can we reject the hypothesis of equality of proportions? We may assume that white and purple plants are equally viable.

It would be natural to argue that, if white and purple flowers were equally likely, the probability of our result would be $1/2^9$. If there is no reason to suspect an excess of white rather than an excess of purple flowers, we must add to this the probability of getting nine purple flowers, which is also $1/2^9$, giving a total probability of $1/2^8$. The hypothesis of equality of proportions would then be rejected on the 1/256, or the 0·3906 % level of significance. But if we did this our reference class would not be the set of indefinite repetitions of the experiment, in its ordinary meaning.

A repetition of the experiment, in its ordinary meaning, would consist of another selection of ten seeds from the bag, and their planting and growth. On such another occasion all ten plants might grow to maturity, or all or some might die. These possibilities have not been taken into account in our calculation of probability, so far.

To allow for the possible variation in the number of plants which grow, we might lay out the set of all possible results of the experiment as in Fig. 1, where n denotes the number of plants that grow, and r denotes the excess of white over purple. Thus any point in the figure can be referred to uniquely by its co-ordinates (n, r). If we now introduce a parameter p, to denote the probability (if it exists) that a plant will grow to maturity, given that it has been selected, the probability associated with the point (n, r) on the hypothesis of equality of proportions of white and purple will be

$$W(n,r;p) = \frac{10!}{n!(10-n)!}p^n(1-p)^{10-n}\frac{n!\,2^{-n}}{(\frac{1}{2}(n+r))!\,(\frac{1}{2}(n-r))!},$$

and since this is a function of the unknown p, we have a special problem of arranging the points (n, r) in order of significance before we can establish a test. The situation in this respect is similar to that dealt with in the paper on 2×2 tables, printed earlier in this issue (Barnard, 1946, pp. 123–38 above).

Proceeding as in the earlier paper, we notice first that the same level of significance must apply to (n, r) as to $(n, -r)$, so that we can confine our further considerations to the upper half of the diagram. Now in this half, the transition from (n, r) to $(n+1, r+1)$ means we discover that one of the plants which failed to grow in our case, was in fact a white-flowered plant. In this case our conviction that there is an excess of white-flowered plants would be strengthened, so that $(n+1, r+1)$ would be reckoned more significant than (n, r). Similarly, going from (n, r) to $(n+1, r-1)$ would mean that a missing plant was found to be purple, and this would weaken our belief in an excess of white-flowered plants; consequently,

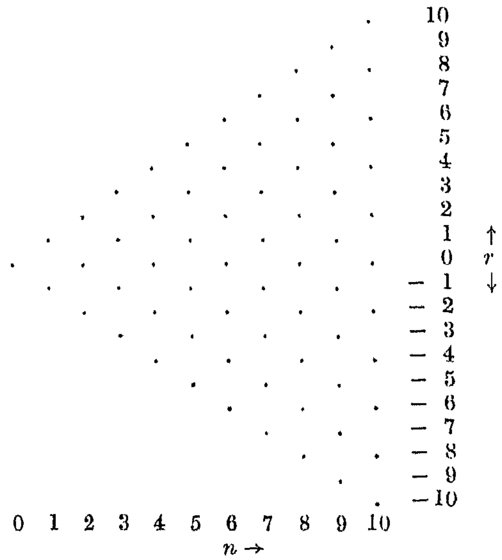


Fig. 1

(n, r) would be reckoned more significant than $(n+1, r-1)$. Finally, going from (n, r) to $(n+2, r)$ would mean growing two more plants, one purple and one white, and this would increase our tendency to believe in the equality of proportions. Consequently, (n, r) would be reckoned more significant than $(n+2, r)$. These principles taken together imply that points lying north-east, or west, of a given point (n, r), or between these two directions, would be reckoned more significant than (n, r); while, conversely, points lying east to south-west (inclusive) from (n, r) would be reckoned less significant than (n, r). The relative significance of points lying inside the half-quadrants north-east to east and south-west to west would remain undetermined.

We could now proceed as in the paper (1), building up a test, consistent with the above partial ordering, in such a way as to make the significance or otherwise of our result depend as little as possible on any knowledge we may have about the value of p. But we need not carry this through for the result we have quoted, since our conditions by themselves require that the only points in the diagram which should be reckoned not less significant than our result are the points (9, 9), (9, −9), (10, 10) and (10, −10). The probability associated with these four points is

$$\begin{aligned} P(9, 9; p) &= 2(10p^9(1-p) \,.\, 2^{-9} + p^{10}2^{-10}) \\ &= (p/2)^9\,(20-19p), \end{aligned}$$

the maximum value of which occurs when $p = 18/19$, and is $P_m(9, 9) = 0{\cdot}002413$. Thus on this basis we should conclude that our result was significant on the $0{\cdot}2413\,\%$ level.

The difference between the first result, 0·3906 %, and the second, 0·2413 %, is in practice negligible. Somewhat larger differences will be found in other similar cases, however, and it seems worth while to try to clarify the cause of the discrepancy.

Consider three possible causes for the failure of the tenth plant to grow to maturity:

(1) The bag from which the seed was taken is known to contain a proportion of dead seeds, which are physically indistinguishable from the live ones, and the tenth seed planted happened to be one of these. The conditions of growth were such that any live seed planted would have grown.

(2) The tenth plant happened to be attacked by a soil pest, which destroyed it.

(3) The statistician trod on the tenth plant while running for a bus; otherwise, it would have grown.

If we now consider what would happen in these three cases if the experiment were repeated, in case (1) we should be just as uncertain as before how many plants would grow, out of those selected. In case (2), we might or might not happen to strike a good year for the pest in question, so that we might or might not have a similar accident recurring. In case (3) we should obviously give the statistician firm instructions not to be careless, and then we could be reasonably certain that all the plants selected would grow.*

In the first case, we can suppose that the proportions of white, purple, and dead seeds in the bag are, respectively, p_1, p_2, and $1-(p_1+p_2)$; and the purpose of our experiment is to test the hypothesis $p_1 = p_2$. In this case, putting $p_1+p_2 = p$, we can clearly apply the analysis of Fig. 1, and the appropriate level of significance is 0·2413 %.

In the third case, the situation actually realized is just what it would have been if we had warned the statistician beforehand, and then thrown one of the ten seeds back into the bag. Thus our effective sample size here is 9, and the appropriate level of significance is 0·3906 %.

In the second case, the answer depends on our attitude to the set of accidents of which the pest is a specimen. If this set of accidents is regarded as a stable set of chance causes we may be justified in representing its effect on the growth of our plants by the probability p. If, on the other hand, the incidence of such pests undergoes, say, regular cyclical fluctuations from year to year, so that its incidence is to some extent predictable, if not wholly controllable, then we should not be justified in assuming the existence of a real probability corresponding to our parameter p. We should, to be on the safe side, in this case allow for the possibility that experimental technique might improve in the future, to such an extent as to eliminate the possibility of such accidents. Thus, adopting this conservative attitude to our results, we should here treat the effective sample size as 9. The repetitions of the experiment which we have in mind would then be imaginary repetitions, in which experimental technique was supposed to be better than it is now, and we have as much control over pests as we have over statisticians.

The general situation illustrated by this example can be described in terms of the notion of 'isolate' introduced by Prof. H. Levy (1931). In making an experiment, we try to construct an isolate—a system, or part of the world, which we suppose has relatively little interaction with the rest of the world, and which, for practical purposes, may be considered on its own. This isolate may contain within itself all the systems of chance causes which are

* It is not suggested that the three cases exhaust the multiplicity of types which might arise in practice. As Prof. Pearson has pointed out, if it were not the statistician, but his three-year-old son who was the vandal in case (3), we should have here a situation intermediate between our second and third instances.

regarded as affecting, to any practical extent, the results of the experiment. Such is the case in (1), where all the chance causes involved in the experiment are supposed given in the bag which is the subject of the experiment. Here, then, we are dealing with a 'good isolate', whose interaction with the rest of the world is really negligible, and chance causes operate within the isolate.

In case (3), on the other hand, we are dealing with an imperfect isolate. The outside world, in the shape of the statistician, interacts with our isolate to an extent not negligible in practice. Fortunately, in this case we are able to construct a smaller isolate, consisting of the nine surviving plants, in which the interactions with the outside world are negligible. In case (2), there may be some doubt as to what isolate we are discussing. If we regard soil pests and such things as included in the isolate, and represent them as a stable set of chance causes, then we are entitled to analyse as in case (1); but if the pests are not included in the isolate, we should analyse as in case (3).

Statistical tests are applicable to at least two types of experiment. First, to experiments in which the isolate studied contains within itself a system of chance causes which may influence the results. And second, to experiments in which the isolate studied is not a 'good' isolate, and the residual interactions with the rest of the world may affect the results. There may also be mixed cases.

The distinction between the two types may also be brought out in relation to the necessity or otherwise of an 'artificial' randomization procedure, using random digits or the like. In the first type, such an artificial randomization procedure is not strictly necessary; for example, with our bag of seeds, the bag itself, and its physically indistinguishable contents, forms a perfectly adequate randomizer. We have in this case, as it were, an impermeable shield around the system, which prevents any external shocks from affecting the system. In the second type of experiment, we need to ensure that the interactions with the outside world will not mask the results we are interested in; and if we cannot ensure a practically complete separation from the outside world, then the effect of external intereactions must be randomized, by a special procedure. The randomization here acts like a shock absorber, specially placed around the experiment to distribute external shocks evenly through the system.

In the first type of experiment, the reference class to which the significance level applies is in fact the set of indefinite repetitions of the experiment in question. In the second type of experiment, the reference class is an ideal set, in which the accidental influences of the outside world repeat themselves exactly, while the effect of these accidents on the system varies as a result of the special randomization.

REFERENCES

BARNARD, G. A. (1946). Significance tests for 2×2 tables. *Biometrika*, **34**, 123.
LEVY, H. (1931). *The Universe of Science*. London: Watts and Co.

Vol. XXXIV. Parts III and IV December, 1947

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

IN CONSULTATION WITH

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

Reprinted by offset-litho 1953

[*Issued* 30 *December*, 1947]

ON THE DISTRIBUTION OF THE RANK CORRELATION COEFFICIENT τ WHEN THE VARIATES ARE NOT INDEPENDENT

BY WASSILY HÖFFDING

I. INTRODUCTION

1. Consider a population distributed according to two variates x, y. Two members (x_1, y_1) and (x_2, y_2) of the population will be called *concordant* if both values of one member are greater than the corresponding values of the other one, that is if

$$x_1 < x_2,\ y_1 < y_2 \quad \text{or} \quad x_1 > x_2,\ y_1 > y_2.$$

They will be called *discordant* if for one member one value is greater and the other one smaller than for the other member, that is if

$$x_1 < x_2,\ y_1 > y_2 \quad \text{or} \quad x_1 > x_2,\ y_1 < y_2.$$

The probability p that two members drawn from the population at random without replacement are concordant will be called the *probability of concordance*, the probability q that they are discordant will be called the *probability of discordance*.

In the following only populations will be considered for which the probabilities of $x_1 = x_2$ or $y_1 = y_2$ are zero, so that

$$p+q = 1. \tag{1}$$

The main types of such populations are (a) an infinite population with both x and y distributed continuously, (b) a finite population where all values of x and all values of y are different among themselves. The condition that the two members are drawn without replacement is, of course, only relevant in case (b).

For a sample of n members drawn from the population, the probabilities of concordance and discordance are defined in the same manner as for the population. They will be denoted by p' and q' to distinguish them from the population values. If for the population (1) is fulfilled, it may be assumed that all values of x and all values of y in the sample are different, so that

$$p'+q' = 1. \tag{2}$$

It follows from the definition that p' is the relative frequency of concordant pairs among the $\binom{n}{2}$ pairs which can be formed from the members of the sample.

The probability of concordance expresses an essential property of a bivariate distribution. It may in itself be considered as a measure of correlation. p' is an estimate of p; it will be shown that the mean value of p' is p. If a coefficient lying between the limits -1 and $+1$ is preferred, the quantity

$$\tau = p'-q' = 2p'-1 \tag{3}$$

may be taken.

2. The quantity p here termed the probability of concordance was apparently first considered by Esscher (1924) who also used the quantity

$$D = \frac{1}{\binom{n}{2}} \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} \operatorname{sign}(x_i - x_j) \operatorname{sign}(y_i - y_j),$$

(where x_i, y_i, $i = 1, \ldots, n$, are the sample values of the variates) which is the same as the coefficient τ as defined by (3). Esscher showed that if x and y are normally correlated with correlation coefficient r, the expectation of $D = \tau$ is

$$E(\tau) = \frac{2}{\pi} \sin^{-1} r. \tag{4}$$

Hence, from this equation, he suggested estimating r from ranked data by means of the relation

$$r = \sin \frac{\pi}{2} \tau = \sin \pi (p' - \tfrac{1}{2}).$$

For the variance of τ Esscher found in the case of a normally distributed population

$$\frac{1}{4} \binom{n}{2} \operatorname{var}(\tau) = pq + \frac{n-2}{2} \left\{ \frac{1}{9} - \left(\frac{2}{\pi} \sin^{-1} \left(\frac{r}{2} \right) \right)^2 \right\}, \tag{5}$$

where
$$4pq = 1 - \left(\frac{2}{\pi} \sin^{-1} r \right)^2.$$

While Esscher saw in p' and $D = \tau$ only a means for estimating r, Lindeberg (1926) stressed the significance of the probability of concordance itself for judging the degree of dependence between the variates. He proposed for that purpose the coefficient

$$P = 100p' - 50 = 50\tau,$$

called by him *Korrelationsprozent*. Lindeberg also gave, without proof, a formula for the variance of p' in the general case of correlated variates (see (13) below).

Jordan (1927) suggested using, instead of Lindeberg's P, the coefficient later termed by Kendall τ.

Kendall (1938), independently of the above authors, proposed τ as a measure of rank correlation. He completely solved the problem of the sampling distribution of τ in a universe in which all possible rankings are equally probable, showing that it rapidly tends to normality for increasing n.

3. The main object of this paper is to show that the sampling distribution of p' (and hence that of τ) tends to normality as $n \to \infty$ for any population with continuously distributed x and y if a certain condition is fulfilled (Part IV). In addition, Lindeberg's formula for the variance of p' is proved (Part II) and extended for a finite population (Part V). Finally, in Part VI the problem of estimating var (p') from the sample is considered.

II. Mean value and variance of p' in the case of an infinite population

4. Consider a sample of n drawn at random from an infinite population with continuous x and y. Replace the values of x and of y in the sample by their ranks and arrange the members of the sample so that the ranking of x is $1, 2, \ldots, n$. Then the ranking of y is a permutation

$$\Pi = (\pi_1, \ldots, \pi_n)$$

of the numbers $1, \ldots, n$.

Let I and J be the numbers of inversions in the permutations $(\pi_n, \pi_{n-1}, \ldots, \pi_1)$ and $(\pi_1, \ldots, \pi_n)$. Then

$$p' = \frac{2I}{n(n-1)}, \quad q' = \frac{2J}{n(n-1)}. \tag{6}$$

Thus the knowledge of the permutation Π corresponding to the given sample is sufficient for evaluating p'.

5. Let $P(\Pi)$ be the probability of drawing a random sample represented by the permutation Π. Let $p'(\Pi)$ be the probability of concordance for such a sample. Then

$$p = \Sigma P(\Pi)\, p'(\Pi), \tag{7}$$

where the sum is extended over all permutations Π of n numbers.

The right-hand side of (7) is equal to the mean value of p'. Hence

$$Ep' = p. \tag{8}$$

Consider, in generalization of p', the probability w' that among $m \leqslant n$ members drawn from the sample at random without replacement, certain pairs of members are concordant; for instance, among four members A, B, C, D, the pairs AB, AC, AD; or the pairs AB, CD, etc. Let w be the corresponding probability for the parent population. Then it is seen in the same manner as with p' that

$$Ew' = w. \tag{9}$$

Thus, if we can express $(p')^\mu$, the probability of drawing μ concordant pairs from the sample, replacing each pair after drawing it, by probabilities without replacement of the type w', we can also, in virtue of (9), represent $E(p')^\mu$ by population parameters of the type w.

6. Now, $(p')^2$, the probability of drawing from the sample one concordant pair and, after replacing it, of drawing again a concordant pair, is the sum of the following three probabilities:

(a) the probability of getting the same pair in both drawings $\left(1\Big/\binom{n}{2}\right)$, multiplied by the probability that this pair is concordant (p');

(b) the probability that the second pair has one member in common with the first pair $\left(2(n-2)\Big/\binom{n}{2}\right)$, multiplied by the probability, say k', that among three members A, B, C drawn from the sample without replacement, one, say A, is concordant with the other two;

(c) the probability that the second pair has no member in common with the first one $\left(\binom{n-2}{2}\Big/\binom{n}{2}\right)$, multiplied by the probability that among four members A, B, C, D drawn without replacement, two pairs without a member in common, say AB and CD, are concordant. The latter probability may be denoted by $(p^2)'$ since the corresponding probability for the infinite population is p^2.

Thus,

$$\binom{n}{2}(p')^2 = p' + 2(n-2)\,k' + \binom{n-2}{2}(p^2)', \tag{10}$$

and, applying (9),

$$\binom{n}{2}E(p')^2 = p + 2(n-2)\,k + \binom{n-2}{2}p^2. \tag{11}$$

Hence, we have for the variance of p'

$$\binom{n}{2}\operatorname{var}(p') = \binom{n}{2}\{E(p')^2 - p^2\} = p + 2(n-2)\,k - (2n-3)\,p^2 \tag{12}$$

or

$$\binom{n}{2}\operatorname{var}(p') = \binom{n}{2}\mu_2(p') = p(1-p) + 2(n-2)\,(k-p^2). \tag{13}$$

This is identical with the formula given without proof by Lindeberg (1926).

7. In the case considered by Kendall where all permutations Π of n numbers are equally probable, the permutations of $m \leqslant n$ also are equiprobable. Hence

$$p = P(1,2) = q = P(2,1) = \tfrac{1}{2}.$$

Further, representing k as the mean value of k' in a sample of 3, we find

$$k = P(123) + \tfrac{1}{3}P(132) + \tfrac{1}{3}P(213) = (1 + \tfrac{1}{3} + \tfrac{1}{3})\frac{1}{3!} = \frac{5}{18}.$$

Inserting these values in (13), we have

$$\operatorname{var}(p') = \frac{2n+5}{18n(n-1)},$$

$$\operatorname{var}(\tau) = 4\operatorname{var}(p') = \frac{2(2n+5)}{9n(n-1)}$$

in accordance with Kendall's formula.

III. Some algebraic formulae

8. We shall now consider some algebraic relations to be used in the proof of normality of p' for large n.

Let $f_d(\rho)$ be a polynomial of degree d in p. Then

$$\sum_{\rho=0}^{\beta} (-1)^{\beta-\rho}\binom{\beta}{\rho} f_d(\rho) = \begin{cases} 0 & \text{if } d < \beta, \\ a_0\beta! & \text{if } d = \beta, \end{cases} \tag{14}$$

where a_0 is the coefficient of the highest power ρ^d in $f_d(\rho)$.

To prove (14) write $\qquad f_d(\rho) = a_0\rho^{[d]} + a_1\rho^{[d-1]} + \dots,$

where $\qquad \rho^{[0]} = 1, \quad \rho^{[\delta]} = \rho(\rho-1)\dots(\rho-\delta+1), \quad (\delta \geqslant 1).$

Then (14) follows from the fact that

$$\sum_{\rho=0}^{\beta} (-1)^{\beta-\rho}\binom{\beta}{\rho}\rho^{[\delta]} = \beta^{[\delta]} \sum_{\rho-\delta=0}^{\beta-\delta} (-1)^{\beta-\rho}\binom{\beta-\delta}{\rho-\delta}$$

is equal to $\beta^{[\delta]}(1-1)^{\beta-\delta} = 0$ if $\beta - \delta > 0$ and to $\beta!$ if $\beta = \delta$.

9. For any non-negative integer ν we may write

$$n^\nu = d_{\nu 0}^{(\alpha)}(n-\alpha)^{[\nu]} + d_{\nu 1}^{(\alpha)}(n-\alpha)^{[\nu-1]} + \dots + d_{\nu,\nu-1}^{(\alpha)}(n-\alpha) + d_{\nu\nu}^{(\alpha)}. \tag{15}$$

We will study certain properties of the coefficients $d_{\nu\kappa}^{(\alpha)}$.

From (15) it is seen immediately that

$$d_{\nu 0}^{(\alpha)} = 1. \tag{16}$$

Inserting in (15) $n = \alpha + \beta$ $(\beta = 0, 1, \dots)$ we have

$$(\alpha+\beta)^\nu = d_{\nu,\nu-\beta}^{(\alpha)}\beta! + d_{\nu,\nu-\beta+1}^{(\alpha)}\beta^{[\beta-1]} + \dots + d_{\nu,\nu-1}^{(\alpha)}\beta + d_{\nu\nu}^{(\alpha)} \tag{17}$$

or

$$d_{\nu\nu}^{(\alpha)} = \alpha^\nu, \quad d_{\nu,\nu-\beta}^{(\alpha)} = \frac{1}{\beta!}\left\{(\alpha+\beta)^\nu - \sum_{\lambda=0}^{\beta-1} \beta^{[\lambda]} d_{\nu,\nu-\lambda}^{(\alpha)}\right\}, \quad (\beta = 1, 2, \dots). \tag{18}$$

Hence we find by induction

$$d_{\nu,\nu-\beta}^{(\alpha)} = \frac{1}{\beta!}\sum_{\rho=0}^{\beta} (-1)^{\beta-\rho}\binom{\beta}{\rho}(\alpha+\rho)^\nu. \tag{19}$$

If we take this as definition of $d_{\nu\kappa}^{(\alpha)}$ for $\kappa < 0$, we have in virtue of (14)

$$d_{\nu\kappa}^{(\alpha)} = 0 \quad \text{for} \quad \kappa < 0. \tag{20}$$

Expanding $(\alpha+\rho)^\nu$ we have from (19)

$$d_{\nu,\nu-\beta}^{(\alpha)} = \frac{1}{\beta!}\sum_{\rho=0}^{\beta} (-1)^{\beta-\rho}\binom{\beta}{\rho}\sum_{\sigma=0}^{\nu}\binom{\nu}{\sigma}\alpha^\sigma\rho^{\nu-\sigma} = \sum_{\sigma=0}^{\nu}\binom{\nu}{\sigma}\alpha^\sigma\frac{1}{\beta!}\sum_{\rho=0}^{\beta}(-1)^{\beta-\rho}\binom{\beta}{\rho}\rho^{\nu-\sigma}.$$

Comparing the last sum with (19) and writing

$$d_{\nu\kappa} = d_{\nu\kappa}^{(0)},$$

we have
$$d_{\nu,\nu-\beta}^{(\alpha)} = \sum_{\sigma=0}^{\nu} \binom{\nu}{\sigma} d_{\nu-\sigma,\nu-\beta-\sigma}\alpha^{\sigma}$$

or, putting $\nu-\beta = \kappa$ and noting that, by (20), $d_{\nu-\sigma,\kappa-\sigma} = 0$ for $\sigma > \kappa$,

$$d_{\nu\kappa}^{(\alpha)} = \sum_{\sigma=0}^{\kappa} \binom{\nu}{\sigma} d_{\nu-\sigma,\kappa-\sigma}\alpha^{\sigma}. \tag{21}$$

We have the recurrence relation

$$d_{\nu+1,\kappa}^{(\alpha)} - d_{\nu\kappa}^{(\alpha)} = (\alpha+\nu+1-\kappa)\, d_{\nu,\kappa-1}^{(\alpha)} \tag{22}$$

which can be obtained by multiplying (15) by n, then writing down (15) with $\nu+1$ instead of ν, and comparing coefficients in both expressions.

10. We prove now two properties of the coefficients $d_{\nu\kappa}^{(\alpha)}$.

(I) $d_{\nu\kappa}$ is a polynomial in ν of degree 2κ, the term of highest degree being $\nu^{2\kappa}/2^{\kappa}\kappa!$.

In virtue of (16) this is true for $\kappa = 0$. And if it is true for $\kappa-1$, the highest term of $d_{\nu+1,\kappa} - d_{\nu\kappa}$ is, by (22) with $\alpha = 0$, $\nu^{2\kappa-1}/2^{\kappa-1}(\kappa-1)!$, and hence that of $d_{\nu\kappa}$, by a well-known theorem,

$$\frac{1}{2\kappa}\frac{1}{2^{\kappa-1}(\kappa-1)!}\nu^{2\kappa} = \frac{1}{2^{\kappa}\kappa!}\nu^{2\kappa}.$$

(II) $d_{t-\rho,\kappa}^{(\gamma-t)}$ is a polynomial in t of degree 2κ with the highest term $(-1)^{\kappa}t^{2\kappa}/2^{\kappa}\kappa!$.

From (21),
$$d_{t-\rho,\kappa}^{(\gamma-t)} = \sum_{\sigma=0}^{\kappa} \binom{t-\rho}{\sigma} d_{t-\rho-\sigma,\kappa-\sigma}(\gamma-t)^{\sigma}.$$

In $\binom{t-\rho}{\sigma}$ the highest term in t is $\frac{1}{\sigma!}t^{\sigma}$.

In $d_{t-\rho-\sigma,\kappa-\sigma}$ the highest term in t is $\frac{1}{2^{\kappa-\sigma}(\kappa-\sigma)!}t^{2\kappa-2\sigma}$ (by (I)).

In $(\gamma-t)^{\sigma}$ the highest term in t is $(-1)^{\sigma}t^{\sigma}$.

Hence, in $d_{t-\rho,\kappa}^{(\gamma-t)}$ the highest term is

$$\sum_{\sigma=0}^{\kappa} \frac{(-2)^{\sigma}}{2^{\kappa}\sigma!(\kappa-\sigma)!}t^{2\kappa} = \frac{1}{2^{\kappa}\kappa!}(1-2)^{\kappa}t^{2\kappa} = \frac{(-1)^{\kappa}}{2^{\kappa}\kappa!}t^{2\kappa}.$$

11. $d_{\nu,\nu-\beta}$ has also a combinatorial meaning.

Let
$$\Sigma_{\beta}(\nu) = \Sigma\frac{\nu!}{\nu_1!\,\nu_2!\ldots\nu_{\beta}!}, \quad \Sigma_{\beta}'(\nu) = \Sigma'\frac{\nu!}{\nu_1!\,\nu_2!\ldots\nu_{\beta}!}$$

where Σ indicates summation over all $\nu_i \geqslant 0$, Σ' over all $\nu_i \geqslant 1$, and in both cases $\nu_1 + \ldots + \nu_{\beta} = \nu$. $\Sigma_{\beta}(\nu)$ is the number of ways of allocating ν objects on β places, and $\Sigma_{\beta}'(\nu)$ is the number of ways of allocating ν objects on β places in such a manner that no place remains empty.

We have
$$\Sigma_{\beta}(\nu) = \beta^{\nu},$$

and a little consideration shows that

$$\Sigma_{\beta}(\nu) = \Sigma_{\beta}'(\nu) + \binom{\beta}{1}\Sigma_{\beta-1}'(\nu) + \ldots + \binom{\beta}{\beta-1}\Sigma_1'(\nu).$$

Comparing this with (17) we see that

$$d_{\nu,\nu-\beta} = \frac{1}{\beta!}\Sigma_{\beta}'(\nu) = \frac{1}{\beta!}\Sigma'\frac{\nu!}{\nu_1!\,\nu_2!\ldots\nu_{\beta}!}. \tag{23}$$

IV. Proof of normality of p' for $n \to \infty$

12. Any set of different pairs of elements belonging to the population will be briefly referred to as a *system* (two pairs being different if they have no more than one element in common).

If we represent the elements of a system by points in a plane and the pairs of elements by lines joining the points, we have a *pattern* corresponding to the given system. Two systems will be said to have the same pattern if there exists a one-to-one correspondence between the elements of both systems such that if two elements of one system form a pair, the two corresponding elements of the other system also form a pair. Thus the only thing relevant in a pattern is the lines connecting the points, the position of the points having no significance.

A pattern will be called *simple* if one can pass from any point of the pattern to any other one along lines belonging to the pattern. A *composite* pattern is a pattern consisting of more than one simple pattern.

If the elements of a system (or the points of the corresponding pattern) are denoted by different letters $A, B, C, \ldots$, each pair of the system can be represented by a pair of letters. All systems of one pair have the same pattern (AB). There are two patterns of two pairs, one simple and containing three points (AB, BC) and one composite and containing four points (AB, CD). There are five patterns of three pairs, three simple $(AB, BC, CA; AB, BC, CD; AB, AC, AD)$, one consisting of two different simple patterns (AB, CD, DE) and one consisting of three equal simple patterns (AB, CD, EF).

13. If a *simple* pattern consists of a_j points and b_j pairs,

$$a_j \leqslant b_j + 1. \tag{24}$$

For this is true for $b_j = 1$, and by adding one pair to a simple pattern, at most one point is added if the new pattern is to be simple again.

Denote the different simple patterns by $S_1, S_2, \ldots, S_j, \ldots$, where S_1 stands for the one-pair pattern and S_2 for the two-pairs pattern (AB, BC), all S_j with $j \geqslant 3$ consisting of three or more pairs. Let a_j be the number of points and b_j the number of pairs in S_j. Then $a_1 = 2$, $a_2 = 3$,

$$b_1 = 1, \quad b_2 = 2, \quad b_j \geqslant 3 \quad \text{if} \quad j \geqslant 3. \tag{25}$$

Consider a pattern P composed of γ_1 simple patterns S_1, γ_2 simple patterns S_2, etc., and containing a points and b pairs. Then, writing symbolically

$$P = \Sigma \gamma_j S_j,$$

we have

$$a = \Sigma \gamma_j a_j, \quad b = \Sigma \gamma_j b_j.$$

In virtue of (24),

$$3b - 2a = \Sigma \gamma_j (3b_j - 2a_j) \geqslant \Sigma \gamma_j (b_j - 2),$$

and from (25)

$$3b - 2a \geqslant -\gamma_1, \tag{26}$$

the sign of equality holding if, and only if, pattern P contains no other simple patterns than S_1 and S_2.

14. $(p')^\mu$ is the probability that μ pairs of elements drawn from the sample, replacing each pair after drawing, are all concordant. We may write

$$(p')^\mu = \Sigma A_i w_i',$$

where A_i is the probability that μ pairs are drawn from the sample in such a way that the system of different pairs among them has the pattern P_i, and if a_i is the number of points in P_i, w_i' is the probability that if a_i elements are drawn from the sample without replacement

and paired according to pattern P_i, all pairs of P_i are concordant. The summation is extended over all patterns P_i with no more than μ pairs.

Since the probabilities w'_i are of the type for which formula (9) is applicable, we have

$$E(p')^\mu = \Sigma A_i w_i, \tag{27}$$

where, as usual, w_i is the population probability corresponding to the sample probability w'_i.

15. Consider a term Aw in (27) corresponding to the pattern

$$P = \Sigma \gamma_j S_j$$

with $\gamma = \Sigma \gamma_j$ simple patterns, $a = \Sigma \gamma_j a_j$ points and $b = \Sigma \gamma_j b_j$ pairs.

Let
$$\overline{P} = \sum_{j \geqslant 2} \gamma_j S_j$$

be *the pattern obtained from* P *by excluding the single-pair patterns* S_1. Then

$$\overline{\gamma} = \gamma - \gamma_1, \quad \overline{a} = a - 2\gamma_1 \quad \text{and} \quad \overline{b} = b - \gamma_1 \tag{28}$$

are the numbers of simple patterns, points and pairs in $\overline{P}$.

We have
$$w = p^{\gamma_1} v, \tag{29}$$

where v is independent of p and γ_1, only depending on the pattern $\overline{P}$.

16. The probability A will be studied, in the first place, as a function of n and γ_1, while its dependence on $\overline{P}$ will be considered later and only in a special case. It must be borne in mind that, by (28), γ, a and b also depend on γ_1.

Let $Q_1, Q_2, \ldots, Q_b$ be the pairs of pattern P numbered in some definite order. Suppose pair Q_β appears μ_β times ($\beta = 1, \ldots, b$). Then

$$\mu_1 + \ldots + \mu_b = \mu, \quad \mu_\beta \geqslant 1 \quad (\beta = 1, \ldots, b).$$

Let $R_1, R_2, \ldots, R_\mu$ be the total set of the pairs drawn, numbered independently of the order in which they appear. Then μ_1 R's are equal to Q_1, μ_2 R's are equal to Q_2, etc.

Let B be the probability that among μ pairs drawn from the sample, replacing each pair after drawing, b pairs are different and arranged according to pattern P, pair Q_β appearing μ_β times ($\beta = 1, \ldots, b$) and the μ pairs being drawn in a definite order, say $R_1, R_2, \ldots, R_\mu$.

Suppose, R_1 is a Q_1. Since any pair drawn may be taken as Q_1 (only the relative position of the pairs being relevant), the probability first to draw R_1 is 1. The probability that the second pair drawn is R_2 depends on whether R_2 has no, one or both elements in common with R_1. In the first case, it is $\binom{n-2}{2}\Big/\binom{n}{2}$, in the second case $2(n-2)\Big/\binom{n}{2}$ (the factor 2 arising from the fact that each of the two elements of R_1 can be the element common with R_2), and in the third case, $1\Big/\binom{n}{2}$.

In general, if the first λ pairs drawn are $R_1, \ldots, R_\lambda$, and if they form a pattern P' containing α different elements, the probability that the $(\lambda+1)$th pair drawn is $R_{\lambda+1}$ depends on whether $R_{\lambda+1}$ has no, one or both elements in common with P'. In the first case it is $\binom{n-\alpha}{2}\Big/\binom{n}{2}$, in the second case, $c'(n-\alpha)\Big/\binom{n}{2}$, and in the third case, $c''\Big/\binom{n}{2}$, where c' and c'' are independent of n. If, in the last case, $R_{\lambda+1}$ is equal to one of the preceding R's, $c'' = 1$.

B is the product of all μ such probabilities, and it is seen from the above consideration that it is of the form

$$B = C(n-2)^{[a-2]}\binom{n}{2}^{-\mu+1},$$

where C is independent of n.

We also see that a pair which has already appeared before makes no contribution to C. Hence, C only depends on the different pairs of pattern P, and is independent of the numbers μ_β.

The above reflexion further shows that for any simple pattern contained in P, the pair drawn first, having no elements in common with the preceding pairs, contributes to C the factor $\frac{1}{2}$, except for the first pair, R_1, which yields the factor 1. Thus, C contains the factor $2^{-\gamma+1} = 2^{-\gamma_1-\bar{\gamma}+1}$, and $2^{-\gamma_1}$ is obviously the sole contribution to C from the γ_1 single pairs (pattern S_1) contained in P. Hence

$$B = 2^{-\gamma_1}C'(n-2)^{[a-2]}\binom{n}{2}^{-\mu+1},$$

where C' is independent of n and γ_1, and also independent of the order in which the γ_1 single pairs are drawn.

A, the probability that μ pairs drawn form pattern P, irrespective of the order in which they appear, depends on n in the same way as B. As a function of γ_1, A contains, besides $2^{-\gamma_1}$, the factor $1/\gamma_1!$ owing to the fact that the γ_1 single pairs are interchangeable. Further it contains the factor $\Sigma_b'(\mu)$ which indicates the number of ways of allocating μ objects on b places so that no place remains empty. In virtue of (23) we have

$$\Sigma_b'(\mu) = b!\,d_{\mu,\mu-b}.$$

Thus, A is of the form
$$A = D^{(\mu)}_{\gamma_1}(n-2)^{[2\gamma_1+\bar{a}-2]}\binom{n}{2}^{-\mu+1}, \tag{30}$$

where
$$D^{(\mu)}_{\gamma_1} = 2^{-\gamma_1}\frac{(\gamma_1+\bar{b})!}{\gamma_1!}d_{\mu,\mu-\gamma_1-\bar{b}}D' \tag{31}$$

and D' is independent of both n and γ_1 and only depends on the pattern $\bar{P}$ containing no S_1.

Inserting (29) and (30) in (27), we have

$$\binom{n}{2}^{\mu-1}E(p')^\mu = \Sigma\, D^{(\mu)}_{\gamma_1}p^{\gamma_1}v(n-2)^{[2\gamma_1+\bar{a}-2]}, \tag{32}$$

the summation taking place over all patterns with no more than μ pairs. (32) also holds for $\mu = 0$ if by a 'pattern of 0 pairs' we understand the case $\gamma_1 = \gamma_2 = \ldots = 0$ and take (31) as definition of $D^{(0)}_{\gamma_1}$ with suitably chosen D'. $\rho^{[-\delta]}$ with $\delta > 0$ is defined by

$$\rho^{[-\delta]}(\rho+\delta)^{[\delta]} = 1.$$

17. If
$$\mu_\nu(p') = E(p'-p)^\nu,$$

we have
$$\binom{n}{2}^{\nu-1}\mu_\nu(p') = \sum_{\delta=0}^{\nu}(-1)^\delta\binom{\nu}{\delta}\binom{n}{2}^{\delta}p^\delta\binom{n}{2}^{\nu-\delta-1}E(p')^{\nu-\delta}.$$

Applying (32) with
$$\mu = \nu-\delta, \quad \gamma_1 = \kappa-\delta, \tag{33}$$

we have for the coefficient of $p^\kappa v$ in $\binom{n}{2}^{\nu-1}\mu_\nu(p')$

$$\sum_{\delta=0}^{\nu}(-1)^\delta\binom{\nu}{\delta}\binom{n}{2}^{\delta}D^{(\nu-\delta)}_{\kappa-\delta}(n-2)^{[\bar{a}+2\kappa-2\delta-2]}$$
$$= \sum_{\delta=0}^{\nu}(-1)^\delta\binom{\nu}{\delta}\frac{1}{2^\delta}\sum_{\rho=0}^{\delta}(-1)^\rho\binom{\delta}{\rho}n^{2\delta-\rho}D^{(\nu-\delta)}_{\kappa-\delta}(n-2)^{[\bar{a}+2\kappa-2\delta-2]}.$$

Inserting here, in accordance with (15),

$$n^{2\delta-\rho} = \sum_{\sigma=0}^{2\delta-\rho} d_{2\delta-\rho,\sigma}^{(\bar{a}+2\kappa-2\delta)} (n-\bar{a}-2\kappa+2\delta)^{[2\delta-\rho-\sigma]},$$

we have $$\sum_{\delta=0}^{\nu} (-1)^{\delta} \binom{\nu}{\delta} \frac{1}{2^{\delta}} \sum_{\rho=0}^{\delta} (-1)^{\rho} \binom{\delta}{\rho} D_{\kappa-\delta}^{(\nu-\delta)} \sum_{\sigma=0}^{2\delta-\rho} d_{2\delta-\rho,\sigma}^{(\bar{a}+2\kappa-2\delta)} (n-2)^{[\bar{a}+2\kappa-2-\rho-\sigma]}.$$

Putting $\alpha = \bar{a}+2\kappa-2-\rho-\sigma$, we have for the coefficient $K_{\kappa}^{(\nu,\alpha)}$ of $p^{\kappa}v(n-2)^{[\alpha]}$ in $\binom{n}{2}^{\nu-1} \mu_{\nu}(p')$

$$K_{\kappa}^{(\nu,\alpha)} = \sum_{\delta=0}^{\nu} (-1)^{\delta} \binom{\nu}{\delta} a_{\kappa}^{(\nu,\alpha)}(\delta), \tag{34}$$

where $$a_{\kappa}^{(\nu,\alpha)}(\delta) = \frac{1}{2^{\delta}} D_{\kappa-\delta}^{(\nu-\delta)} \sum_{\rho=0}^{\delta} (-1)^{\rho} \binom{\delta}{\rho} d_{2\delta-\rho,\bar{a}+2\kappa-\alpha-\rho-2}^{(\bar{a}+2\kappa-2\delta)}. \tag{35}$$

Since $d_{2\delta-\rho,\bar{a}+2\kappa-\alpha-\rho-2}^{(\bar{a}+2\kappa-2\delta)} = 0$ if $\bar{a}+2\kappa-\alpha-\rho-2<0$, the upper limit of ρ in the summation may be taken as $\bar{a}+2\kappa-\alpha-2$, which is independent of δ. We have then, in virtue of (31),

$$a_{\kappa}^{(\nu,\alpha)}(\delta) = \frac{1}{2^{\kappa}} (\bar{b}+\kappa-\delta)^{[\bar{b}]} d_{\nu-\delta,\nu-\bar{b}-\kappa} D' \sum_{\rho=0}^{\bar{a}+2\kappa-\alpha-2} (-1)^{\rho} \binom{\delta}{\rho} d_{2\delta-\rho,\bar{a}+2\kappa-\alpha-\rho-2}^{(\bar{a}+2\kappa-2\delta)}, \tag{36}$$

where D' is independent of δ.

In virtue of (I) and (II), para. 10, $a_{\kappa}^{(\nu,\alpha)}(\delta)$ is a polynomial in δ. The degree of the $(\rho+1)$th term in the sum in (36) is $\rho+2(\bar{a}+2\kappa-\alpha-\rho-2)$, which is highest for $\rho=0$. Hence, the degree d of $a_{\kappa}^{(\nu,\alpha)}(\delta)$ is

$$d = \bar{b}+2(\nu-\bar{b}-\kappa)+2(\bar{a}+2\kappa-\alpha-2) = 2(\nu+\bar{a}+\kappa-\alpha-2)-\bar{b}. \tag{37}$$

Now, according to (34) and (14), $K_{\kappa}^{(\nu,\alpha)} = 0$ if $d<\nu$, or, in virtue of (37),

$$K_{\kappa}^{(\nu,\alpha)} = 0 \quad \text{if} \quad 2\alpha > \nu-4+2\bar{a}-\bar{b}+2\kappa. \tag{38}$$

Applying (26) for pattern $\bar{P}$, we have, since $\gamma_1 = 0$, $2\bar{a} \leqslant 3\bar{b}$, and consequently

$$2\bar{a}-\bar{b}+2\kappa \leqslant 2(\bar{b}+\kappa),$$

the sign of equality holding if and only if pattern P contains no other simple patterns than S_1 and S_2.

Remembering that, according to para. 16, $b \leqslant \mu$, we have in virtue of (33)

$$\bar{b}+\kappa = \bar{b}+\gamma_1+\delta = b+\delta \leqslant \mu+\delta = \nu. \tag{39}$$

Thus in any case $$2\bar{a}-\bar{b}+2\kappa \leqslant 2\nu$$

and, in virtue of (38), $$K_{\kappa}^{(\nu,\alpha)} = 0 \quad \text{if} \quad \alpha \geqslant \tfrac{3}{2}\nu-1. \tag{40}$$

If P contains at least one simple pattern with more than two pairs, we even have

$$2\bar{a}-\bar{b}+2\kappa < 2\nu,$$

and consequently $$K_{\kappa}^{(\nu,\alpha)} = 0 \quad \text{if} \quad \alpha \geqslant \tfrac{3}{2}\nu-2. \tag{41}$$

From (40) it appears that the degree in n of $\binom{n}{2}^{\nu-1} \mu_{\nu}(p')$ is

$$\leqslant 3h-2 = \tfrac{3}{2}\nu-2 \quad \text{if} \quad \nu = 2h,$$

$$\leqslant 3h-1 = \tfrac{3}{2}\nu-\tfrac{5}{2} \quad \text{if} \quad \nu = 2h+1.$$

Thus, in $\mu_{2h+1}(p')$, if expanded in powers of n, the degree of the highest term is

$$\leqslant 3h-1-4h = -h-1.$$

In $\mu_2(p')$, in virtue of (13), the degree of the highest term is -1, provided that $k-p^2 \neq 0$. Hence, the degree of

$$\alpha_{2h+1}(p') = \frac{\mu_{2h+1}(p')}{\mu_2^{\frac{1}{2}(2h+1)}(p')}$$

is $\leqslant -h-1+h+\frac{1}{2} = -\frac{1}{2}$. It follows that

$$\alpha_{2h+1}(p') \underset{n}{\rightarrow} 0 \quad \text{if} \quad k-p^2 > 0. \tag{42}$$

($k-p^2 < 0$ is impossible since in this case var (p') would become < 0 for large n.)

18. As we have seen, we may write

$$\binom{n}{2}^{2h-1} \mu_{2h}(p') = R_h(n-2)^{[3h-2]} + R_h'(n-2)^{[3h-3]} + \ldots.$$

Then it follows from (41) that R_h only contains terms depending on patterns S_1 and S_2, that is, R_h is of the form

$$R_h = \sum_{\kappa} \sum_{\lambda} K_{\kappa,\lambda}^{(2h,3h-2)} p^\kappa k^\lambda. \tag{43}$$

The only terms in $\binom{n}{2}^{\mu-1} E(p')^\mu$ which can contribute to this sum are of the form

$$D_{\gamma_1}^{(\mu)} p^{\gamma_1} k^\lambda (n-2)^{[2\gamma_1+3\lambda-2]}.$$

The pattern corresponding to such a term is

$$P = \gamma_1 S_1 + \lambda S_2.$$

Remembering the considerations in para. 16, we see that in each S_2 the pair drawn first contributes to C the factor $\frac{1}{2}$ (except if it is R_1), while the first drawing of the other pair yields the factor 2. Hence, the S_2's make no contribution to C, and we have

$$C = 2^{-\gamma_1+1}.$$

The contribution of the patterns S_2 to A is twofold: since in each S_2 the two pairs may be interchanged, this gives the factor $(1/2!)^\lambda$; and since the λ patterns S_2 may be interchanged, we have the factor $1/\lambda!$. Thus

$$D_{\gamma_1}^{(\mu)} = 2^{-\gamma_1-\lambda+1} \frac{(\gamma_1+2\lambda)!}{\gamma_1!\,\lambda!} d_{\mu,\mu-\gamma_1-2\lambda}$$

and, in virtue of (31), since $\bar{b} = 2\lambda$,

$$D' = \frac{1}{2^{\lambda-1}\lambda!}. \tag{44}$$

Inserting in (38) $\quad \nu = 2h, \quad \alpha = 3h-2, \quad \bar{a} = 3\lambda, \quad \bar{b} = 2\lambda, \qquad (45)$

we see that $K_{\kappa,\lambda}^{(2h,3h-2)} \neq 0$ is possible only if

$$\kappa + 2\lambda \geqslant 2h.$$

On the other hand, from (39), $\quad \kappa + 2\lambda \leqslant 2h.$

Hence, $\quad \kappa = 2h - 2\lambda. \qquad (46)$

Inserting this in (43), we have

$$R_h = \sum_{\lambda=0}^{h} K_{2(h-\lambda),\lambda}^{(2h,3h-2)} p^{2(h-\lambda)} k^\lambda, \tag{47}$$

where $\quad K_{2(h-\lambda),\lambda}^{(2h,3h-2)} = \sum_{\delta=0}^{2h} (-1)^\delta \binom{2h}{\delta} a_{2(h-\lambda)}^{(2h,3h-2)}(\delta).$

According to (37) in connexion with (45), (46), the degree in δ of $a^{(2h,3h-2)}_{2(h-\lambda)}(\delta)$ is $2h$. The highest term, $a_0\delta^{2h}$, is contained in the term corresponding to $\rho=0$ in (36). Inserting in (36) the values from (44), (45) and (46) and putting in the sum $\rho=0$, we have

$$2^{-2h+\lambda+1}\frac{(2h-\delta)^{[2\lambda]}}{\lambda!}d_{2h-\delta,0}\,d^{(4h-\lambda-2\delta)}_{2\delta,h-\lambda}.$$

Thus, in virtue of (16) and (II), para. 10,

$$a_0=2^{-2h+\lambda+1}\frac{1}{\lambda!}(-1)^{h-\lambda}\frac{2^{2h-2\lambda}}{2^{h-\lambda}(h-\lambda)!}=\frac{(-1)^{h-\lambda}}{2^{h-1}h!}\binom{h}{\lambda}.$$

According to (14), $$K^{(2h,3h-2)}_{2(h-\lambda),\lambda}=\frac{(-1)^{h-\lambda}(2h)!}{2^{h-1}h!}\binom{h}{\lambda}.$$

Inserting this in (47) we have

$$R_h=\frac{(2h)!}{2^{h-1}h!}\sum_{\lambda=0}^{h}(-1)^{h-\lambda}\binom{h}{\lambda}p^{2(h-\lambda)}k^{\lambda}=\frac{(2h)!}{2^{h-1}h!}(k-p^2)^h.$$

The highest term of $\mu_{2h}(p')$ is thus

$$2^h\frac{(2h)!}{h!}(k-p^2)^h n^{-h},$$

that of $\mu_2(p')$ is $4(k-p^2)n^{-1}$, and hence

$$\alpha_{2h}(p')=\frac{\mu_{2h}(p')}{\mu_2^h(p')}\underset{n}{\rightarrow}\frac{(2h)!}{2^h h!}\quad\text{if}\quad k-p^2>0. \tag{48}$$

From (42) and (48) it follows according to the Second Limit Theorem that the distribution of p' tends to normality as $n\to\infty$, provided that the marginal distributions are continuous and $k-p^2>0$.

The condition $k-p^2>0$ is fulfilled if the population is distributed normally. For, comparing Esscher's formula (5) with (13), we find, since $\operatorname{var}(\tau)=4\operatorname{var}(p')$,

$$k-p^2=\frac{1}{36}-\left\{\frac{1}{\pi}\sin^{-1}\left(\frac{r}{2}\right)\right\}^2.$$

The right-hand side is positive if $|r|<1$.

V. THE VARIANCE OF p' IN THE CASE OF A FINITE POPULATION

19. Consider a sample of n drawn from a finite population of N in which all values of x and all values of y are different. For the sample probabilities p', k', $(p^2)'$, ... we write now

$$p^{(n)},\quad k^{(n)},\quad (p^2)^{(n)},\quad \ldots,$$

and for the corresponding population probabilities

$$p^{(N)},\quad k^{(N)},\quad (p^2)^{(N)},\quad \ldots.$$

Equation (9) remains valid and may be written as follows:

$$Ew^{(n)}=w^{(N)}. \tag{49}$$

In particular, $$Ep^{(n)}=p^{(N)}.$$

The essential difference between this case and the case $N=\infty$ considered above is that the composite probabilities such as $(p^2)^{(N)}$ or $(pk)^{(N)}$ are not equal to $(p^{(N)})^2$ or $p^{(N)}k^{(N)}$. For

instance, $(p^{(N)})^2$ is evidently the same function of $p^{(N)}$, $k^{(N)}$, $(p^2)^{(N)}$ and N as $(p')^2$ is of p', k', $(p^2)'$ and n. Thus we have, replacing n by N in (10),

$$(p^{(N)})^2 = \frac{2}{N^{[2]}}p^{(N)} + \frac{4(N-2)}{N^{[2]}}k^{(N)} + \frac{(N-2)^{[2]}}{N^{[2]}}(p^2)^{(N)}, \tag{50}$$

and hence
$$(p^2)^{(N)} = \frac{N^{[2]}}{(N-2)^{[2]}}(p^{(N)})^2 - \frac{4}{N-3}k^{(N)} - \frac{2}{(N-2)^{[2]}}p^{(N)}. \tag{51}$$

On the other hand, from (10) and (49)

$$E(p^{(n)})^2 = \frac{2}{n^{[2]}}p^{(N)} + \frac{4(n-2)}{n^{[2]}}k^{(N)} + \frac{(n-2)^{[2]}}{n^{[2]}}(p^2)^{(N)}, \tag{52}$$

which is the equivalent of equation (11).

On subtracting (50) from (52) we find

$$\operatorname{var}(p^{(n)}) = \frac{2(N-n)}{n(n-1)N(N-1)}\{(N+n-1)p^{(N)} + 2[Nn-2(N+n-1)]k^{(N)} - [2Nn-3(N+n-1)](p^2)^{(N)}\}. \tag{53}$$

Substituting for $(p^2)^{(N)}$, the expression in (51), we obtain

$$\operatorname{var}(p^{(n)}) = \frac{2(N-n)}{n(n-1)(N-2)(N-3)}\{(N+n-5)p^{(N)}(1-p^{(N)}) + 2(n-2)(N-2)(k^{(N)}-p^{(N)2})\}, \tag{54}$$

or
$$\binom{n}{2}\operatorname{var}(p^{(n)}) = \left(1-\frac{(n-2)(n-3)}{(N-2)(N-3)}\right)p^{(N)}(1-p^{(N)}) + 2(n-2)\left(1-\frac{n-3}{N-3}\right)(k^{(N)}-p^{(N)2}). \tag{55}$$

For $N\to\infty$, (55) becomes the same as (13).

VI. A sample estimate of var (p')

20. *In the case of an infinite population, let*

$$\binom{n}{2}\operatorname{var}'(p') = p' + 2(n-2)k' - (2n-3)(p^2)'. \tag{56}$$

Then, in virtue of (9) and (12),

$$E\operatorname{var}'(p') = \operatorname{var}(p').$$

On inserting in (56) for $(p^2)'$ the expression obtained from (10), we find

$$\binom{n-2}{2}\operatorname{var}'(p') = p'(1-p') + 2(n-2)(k'-(p')^2),$$

or
$$\operatorname{var}'(p') = \frac{2}{(n-2)(n-3)}p'q' + \frac{4}{n-3}(k'-(p')^2). \tag{57}$$

By analogy,
$$\operatorname{var}'(q') = \frac{2}{(n-2)(n-3)}p'q' + \frac{4}{n-3}(l'-(q')^2), \tag{58}$$

where l' is the probability that among three members A, B, C drawn from the sample without replacement, one, say A, is discordant with the other two.

In the case of a finite population of the type considered in para. 19, we define in a similar way a statistic $\operatorname{var}^{(n)}(p^{(n)})$ such that

$$E\operatorname{var}^{(n)}(p^{(n)}) = \operatorname{var}(p^{(n)}).$$

We find

$$\operatorname{var}^{(n)}(p^{(n)}) = \frac{2(N-n)}{N(N-1)(n-2)(n-3)}\{(N+n-5)\,p^{(n)}(1-p^{(n)})+2(n-2)(N-2)(k^{(n)}-p^{(n)2})\}. \tag{59}$$

A comparison between (59) and (54) shows that $\operatorname{var}^{(n)}(p^{(n)})$ is obtained from $\operatorname{var}(p^{(n)})$ by interchanging n and N and taking the opposite sign.

21. Let g_ν and h_ν be the numbers of sample members concordant and discordant with $A_\nu = (x_\nu, y_\nu)$ $(\nu = 1, \ldots, n)$. The probability of drawing first the member A_ν, and then, without replacing it, a member concordant with A_ν is $\frac{1}{n}\frac{g_\nu}{n-1}$. The probability of drawing, without replacement, first A_ν and then two other members concordant with A_ν is $\frac{1}{n}\frac{g_\nu(g_\nu-1)}{(n-1)(n-2)}$. Hence

$$p' = \frac{\Sigma g_\nu}{n(n-1)}, \quad k' = \frac{\Sigma g_\nu^2 - \Sigma g_\nu}{n(n-1)(n-2)}. \tag{60}$$

Similarly,

$$q' = \frac{\Sigma h_\nu}{n(n-1)}, \quad l' = \frac{\Sigma h_\nu^2 - \Sigma h_\nu}{n(n-1)(n-2)}. \tag{61}$$

If only the value of p' or q' is required, the use of (6) may be more expedient than that of (60) or (61). If, however, the variance, and hence k' or l', is wanted, the calculation by means of the numbers g_ν and h_ν (whose sums are twice the numbers of inversions I, J) according to (60) or (61) is to be preferred.

If $p' > \frac{1}{2}$, it is more convenient to calculate q' and l' from (61); if $p' < \frac{1}{2}$, the calculation of p' and k' by (60) is more rapid. In many cases one can see directly from the given data whether the concordant or the discordant pairs prevail, before actually calculating p' or q'.

Since $p' + q' = 1$, we have

$$\operatorname{var}(p') = \operatorname{var}(q'),$$

and also, in the case of a finite population,

$$\operatorname{var}(p^{(n)}) = \operatorname{var}(q^{(n)}).$$

If we write down the equation for $\operatorname{var}(q^{(n)})$ analogous to (55) and subtract it from (55), we have

$$k^{(N)} - p^{(N)2} - l^{(N)} + q^{(N)2} = 0,$$

or

$$k^{(N)} - l^{(N)} = p^{(N)2} - q^{(N)2} = p^{(N)} - q^{(N)}.$$

Substituting n for N, we have

$$k' - l' = p' - q' = \tau.$$

Comparing this with (57) and (58) we see that

$$\operatorname{var}'(p') = \operatorname{var}'(q').$$

REFERENCES

ESSCHER, F. (1924). On a method of determining correlation from the ranks of the variates. *Skand. Aktuar.* 7, 201–19.

JORDAN, CH. (1927). *Statistique mathématique*. Paris: Gauthiers-Villars.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30, 81–93.

LINDEBERG, J. W. (1926). Ueber die Korrelation. *Den VI skandinaviske Matematikerkongres i København*, 31 August–4 September 1925, pp. 437–46. København: J. Gjellerup.

ADDENDUM

On p. 184 above, I quoted J. W. Lindeberg as having given the formula for the variance of the probability of concordance p' without proof. I was not aware then that a proof of this formula, as well as that of the corresponding expression for a finite population (equation (54) of my paper), is contained in another paper by Lindeberg, 'Some remarks on the mean error of the percentage of correlation,' *Nordic Statistical Journal*, 1, 137–41 (1929).

THE SIGNIFICANCE OF RANK CORRELATIONS WHERE PARENTAL CORRELATION EXISTS

By H. E. DANIELS (*Wool Industries Research Association*)
and M. G. KENDALL

1. All the known tests of significance of rank correlation coefficients are based on distributions from a population in which each possible ranking occurs equally frequently, i.e. the null case where no parental correlation exists. We may then say of any particular coefficient whether it is significant in the sense that it cannot have arisen with any acceptable probability from an uncorrelated population. No tests are known in the case where parental correlation exists, and we have not seen the point discussed except in reference to the replacement of rank correlations by grade or product-moment correlations. Thus, for example, if two rank correlation coefficients are both found to be significant there has hitherto been no exact method of deciding whether their difference is significant. In this paper we consider the problem of determining confidence intervals for a rank correlation when the parent is correlated and develop a test of significance for the difference of two correlations.

2. In testing an ordinary product-moment correlation the problem is enormously simplified by the assumption that the population is normal, or the further assumption that normal theory holds good even when the parent deviates only moderately from normality. Apart from means and variances the population is then completely specified by the single parent parameter ρ and, as is well known, the sample distribution of the estimator depends only on ρ and the sample number n.

In ranking theory this position no longer obtains. No assumption can in general be made about the form of the parent distribution and, in particular, the parent correlation does not completely specify the problem. The usual type of variate theory cannot, therefore, be expected to meet the requirements.

3. A satisfactory approach to the problem can, however, be made if the rank correlation is measured by the coefficient known as τ (Kendall, 1943, chap. 16). We shall then show that, for large samples at any rate, the problem admits of a solution.

Let the population consist of N members. They may be imagined as laid out in the natural order $1, 2, \ldots, N$ according to the first variate. The rankings according to the second variate are then some permutation of the numbers 1 to N, and this second array of ranks is all we need write down in particular cases. It determines the rank correlation τ. Now suppose we choose a sample of n in one of the $\binom{N}{n}$ possible ways. This sample will, so far as the first variate is concerned, be in the natural order, and the ranks according to the second variate permit of the calculation of a sample correlation t. For all possible samples and any given arrangement of the parent members there will be a distribution of $\binom{N}{n}$ values of t.

4. The sample value of t is an unbiased estimator of τ; that is to say, the mean value of t in all possible samples is τ. For consider the $\binom{N}{n}$ samples of n. Any particular pair of members will occur in $\binom{N-2}{n-2}$ samples, that is, all pairs occur equally frequently in the totality of all samples. In calculating t we assign to any pair $+1$ if its members are in the right order and -1 in the contrary case. Thus the total of the score for all samples is $\binom{N-2}{n-2}$ times the score for the population. To obtain t we divide the score for any sample by $\frac{1}{2}n(n-1)$, and to obtain τ we divide the population score by $\frac{1}{2}N(N-1)$. Hence if Σ is the score for the population, the mean value (expectation) of t is

$$E(t) = \frac{\binom{N-2}{n-2}\Sigma}{\frac{1}{2}n(n-1)\binom{N}{n}} = \frac{\Sigma}{\frac{1}{2}N(N-1)} = \tau. \tag{4·1}$$

5. Unfortunately, it is not true that higher moments of t depend only on τ. A single example will illustrate the point. Consider the ranking of 9:

$$5 \quad 2 \quad 3 \quad 1 \quad 6 \quad 7 \quad 8 \quad 9 \quad 4.$$

If the $84 = \binom{9}{3}$ possible samples of three are written down and t evaluated for each, the distribution of S (the number of positive pairs) is found to be as follows:

Values of S	Frequency
0	2
1	15
2	34
3	33
Total	84

The mean of this distribution is $182/84 = 13/6$, and since

$$t = \frac{2S}{\frac{1}{2}n(n-1)} - 1,$$

the mean value of t is $(26/18) - 1 = 0·44$. The value of S for the parent ranking is 26 and hence $\tau = (52/36) - 1 = 0·44$, verifying equation (4·1). The ranking

$$1 \quad 2 \quad 5 \quad 9 \quad 3 \quad 6 \quad 7 \quad 8 \quad 4$$

also has $\tau = 0·44$, but the distribution of S in samples of three is now:

Values of S	Frequency
0	3
1	16
2	29
3	36
Total	84

The second moment of this distribution is 5·429, against 5·333 for the first distribution, the variances being 0·734 against 0·639.

6. Thus for any parent with given τ there is in general more than one sampling distribution of t according to the arrangement of the parent ranks. In short, as mentioned above, the parameter τ does not completely specify the sampling distribution and in asking the question: What is the standard error of t? we are seeking for an answer which does not exist.

It will be shown, however, that for any given parent ranking the distribution of t tends to normality with increasing n. The sampling properties of t can therefore be specified to a first approximation by its first and second moments only, when the samples are not too small. Further, it will be proved that for given τ the variance of t cannot exceed a certain function of τ and n whatever the parent ranking. From a knowledge of t and n only, it is thus possible to set outer bounds to confidence intervals for τ provided n is large enough for the normal approximation to hold. The limits obtained in this way are sometimes rather wide, and an alternative procedure is to estimate the true variance of t directly from the sample itself according to a formula given below. This avoids the loss of efficiency consequent on using an upper limit to the variance, but it is not known how large a sample is required for the error of estimation to be tolerable.

7. The development of the theory is facilitated if we introduce at the present stage a notation similar to that used by Daniels (1944). The ith and jth ranks corresponding to the second variate are together assigned a score a_{ij} which takes the value $+1$ if the members are in the correct order, -1 if in the wrong order, and a_{ii} is defined to be zero. The ranks for the first variate are similarly assigned scores b_{ij}, but as the members have been taken in the correct order for this variate, the scores are simply $b_{ij} = \pm 1$, $i \lessgtr j$; $b_{ii} = 0$. Next we define $c_{ij} = a_{ij}b_{ij}$, so that $c_{ij} = \pm 1$, according to whether the ranks for the two variates agree or differ in order, and $c_{ii} = 0$. In this notation

$$\tau = c/N(N-1),$$

where $c = \Sigma c_{ij}$, i and j both being summed from 1 to N.

When the sample of n pairs is selected at random from the parent N and its coefficient t is calculated, the values of c_{ij} for the members of the sample remain the same as in the population. This fact makes τ much more suitable for the present problem than the Spearman coefficient ρ whose associated scores do not possess the same property. The sample rank correlation is then

$$t = c^{(n)}/n(n-1),$$

where $c^{(n)} = \Sigma^{(n)} c_{ij}$ and $\Sigma^{(n)}$ denotes summation only over those values of i and j occurring in the sample.

8. It has already been proved that $E(t) = \tau$. To find the variance of t we require $E(t^2)$, so consider

$$\sum_n [c^{(n)}]^2 = \sum_n \Sigma^{(n)} c_{ij} c_{kl},$$

$\sum\limits_n$ denoting summation over all selections of the sample of n from the finite parent population of N members. Let us enumerate the number of ways in which $c_{ij}c_{kl}$ and similar products with 'tied' suffixes, such as $c_{ij}c_{il}$, occur in the sum.

(i) When i, j, k, l are all different the term $c_{ij}c_{kl}$ may occur with $\binom{N-4}{n-4}$ selections of the remaining members of the sample and the contribution of such terms to $\sum_n$ is $\binom{N-4}{n-4}\ \Sigma' c_{ij}c_{kl}$, Σ' meaning summation over all unequal values of i, j, k, l from 1 to N.

(ii) The term $c_{ij}c_{il}$ similarly occurs in $\binom{N-3}{n-3}$ ways and there are four ways of tying one suffix, each of which gives the same contribution to Σ since c_{ij} is symmetrical. The total contribution of such terms to $\sum_n$ is therefore $4\binom{N-3}{n-3}\overset{n}{\Sigma'} c_{ij}c_{il}$.

(iii) Terms like $c_{ij}c_{ij}$ similarly contribute $2\binom{N-2}{n-2}\ \Sigma' c_{ij}c_{ij}$ to $\sum_u$, and all other terms are zero since $c_{ii} = 0$. Hence

$$\sum_n [c^{(n)}]^2 = \binom{N-4}{n-4}\Sigma' c_{ij}c_{kl} + 4\binom{N-3}{n-3}\Sigma' c_{ij}c_{ik} + 2\binom{N-2}{n-2}\Sigma' c_{ij}c_{ij}.$$

Expressing the Σ''s in terms of the corresponding Σ's and dividing out by $\binom{N}{n}$ we obtain

$$E[c^{(n)}]^2 = \frac{n^{(4)}}{N^{(4)}}(\Sigma c_{ij}c_{kl} - 4\Sigma c_{ij}c_{il} + 2\Sigma c_{ij}c_{ij}) + \frac{4n^{(3)}}{N^{(3)}}(\Sigma c_{ij}c_{il} - \Sigma c_{ij}c_{ij}) + \frac{2n^{(2)}}{N^{(2)}}\Sigma c_{ij}c_{ij},$$

where $n^{(r)} = n(n-1)\dots(n-r+1)$. Since $\Sigma c_{ij}c_{ij} = N(N-1)$ and $\Sigma c_{ij}c_{kl} = c^2$, the variance of t for given τ and n is seen to depend on the value of $\Sigma c_{ij}c_{ik} = \Sigma c_i^2$, where $c_i = \sum_{j=1}^{N} c_{ij}$.

Let N become large. The quantities c and Σc_i^2 are respectively $O(N^2)$ and $O(N^3)$, so if we introduce $\tau_i = c_i/N$ the value of $E(t^2)$ for large N becomes

$$E(t^2) \sim \frac{(n-2)(n-3)}{n(n-1)}\tau^2 + \frac{4(n-2)}{n(n-1)}\frac{\Sigma\tau_i^2}{N} + \frac{2}{n(n-1)},$$

and hence in the limit the variance of t is

$$\operatorname{var} t = \frac{4(n-2)}{n(n-1)}\operatorname{var}\tau_i + \frac{2}{n(n-1)}(1-\tau^2). \tag{8·1}$$

9. The variance of t satisfies the inequality

$$\operatorname{var} t \leqslant \frac{2}{n}(1-\tau^2), \tag{9·1}$$

whatever the parent ranking. Moreover, though the limit may not be attained in any particular parent ranking, reasons are given in the Appendix for expecting that it cannot be substantially improved upon. The proof is as follows.

Reverting to a finite parent population of N members, we first seek a maximum for Σc_i^2. In terms of the original scores, $c_{ij} = a_{ij}b_{ij}$. Keeping $b_{ij} = \pm 1$, $i \lessgtr j$, $b_{ii} = 0$, as before, allow the a_{ij}'s to assume *any* values subject to the conditions

$$\Sigma a_{ij}^2 = N(N-1), \quad \Sigma a_{ij}b_{ij} = c = N(N-1)\tau.$$

The stationary values of Σc_i^2 occur when the a_{ij}'s satisfy the equations

$$b_{ij}(c_i + c_j) - \lambda a_{ij} - \mu b_{ij} = 0,$$

which give, on multiplying by b_{ij} and summing j,

$$c_i = \frac{\mu(N-1)-c}{(N-2-\lambda)}.$$

Thus, unless the c_i's are all to be equal, in which case Σc_i^2 is a minimum, λ and μ must take the values

$$\lambda = N-2, \quad \mu = c/(N-1),$$

and since

$$2\Sigma c_i^2 - \lambda N(N-1) - \mu c = 0,$$

it follows that Σc_i^2 cannot exceed $\frac{1}{2}N(N-1)(N-2)+\frac{1}{2}c^2/(N-1)$. Allowing N to become large, this implies

$$\Sigma \tau_i^2/N \leqslant \tfrac{1}{2}(1+\tau^2).$$

Hence

$$\operatorname{var} \tau_i \leqslant \tfrac{1}{2}(1-\tau^2),$$

and so from equation (8·1)

$$\operatorname{var} t \leqslant \frac{2}{n}(1-\tau^2). \tag{9·1}$$

10. Assuming that the sample is large enough for the distribution of t to be normal, the roots τ_1, τ_2 of the equation

$$t-\tau = x\sqrt{\left[\frac{2}{n}(1-\tau^2)\right]}, \tag{10·1}$$

i.e.

$$\tau = \frac{t \pm x\sqrt{\frac{2}{n}}\sqrt{\left(1+\frac{2x^2}{n}-t^2\right)}}{\left(1+\frac{2x^2}{n}\right)}, \tag{10·2}$$

provide confidence limits to τ when t is known, x being the standardized normal deviate corresponding to a given probability of P %. These confidence limits are of course maxima, in the sense that we shall be wrong in *at most* P % of the cases in asserting τ to lie between the calculated limits.

In our proof of the tendency of t to normality it will be necessary to neglect terms of order $n^{-\frac{1}{2}}$, and the sample may have to be rather large for such terms to be small, unless τ itself is small.

The form of equation (9·1) suggests using

$$w = \sin^{-1} t$$

instead of t. To the same order of approximation we can take w as having a normal distribution with mean $\omega = \sin^{-1}\tau$ and standard error not exceeding $\sqrt{(2/n)}$, which is independent of τ. This form is more convenient for assigning confidence limits to t, and for testing the significance of the difference between t_1 and t_2 (whose standard error cannot exceed $\sqrt{[2(1/n_1+1/n_2)]}$), but we have not been able to discover whether the transformation brings the distribution nearer to normality.

11. We now prove that the distribution of t tends for large n to normality whatever the parent ranking, provided that $|\tau|$ is not near unity.

Write $g_{ij} = c_{ij} - c/N^2$ so that $\Sigma g_{ij} = 0$, $g_{ij} = g_{ji}$ and $g_{ii} = -c/N^2 = -(N-1)\tau/N$. The rth moment of $c^{(n)}$ about its mean value is $E[\Sigma^{(n)} g_{ij}]^r$, so consider

$$\sum_n [\Sigma^{(n)} g_{ij}]^r = \sum_n \Sigma^{(n)} g_{ij} g_{kl} g_{uv} \cdots,$$

the summation $\sum\limits_n$ being over all possible sample selections.

The argument used by Daniels (1944) to show that in the null case the distribution of rank correlation in large samples tends to normality can be applied with little modification to the present problem. The proof is therefore sketched here without much detail.

Two essential conditions to be satisfied are that $\Sigma g_{ij} = 0$, which is true by definition, and $\Sigma g_{ij} g_{ik} = O(N^3)$, which is true only if $1-\tau^2 = O(1)$, so that the tendency to normality may be expected to break down for high correlations.

The sum $\underset{n}{\Sigma}$ is evaluated as in § 8 by counting the number of ways in which terms like $g_{ij} g_{kl} g_{uv} \ldots$, and similar terms with tied suffixes, occur. In this way it is expressed as a linear combination of $\Sigma' g_{ij} g_{kl} g_{uv} \ldots$, etc. Every such Σ' is replaceable by the corresponding Σ together with terms containing more tied suffixes which are of lower order in N since they involve fewer summations from 1 to N.

12. First consider the even moments with $r = 2m$. Terms containing more than $3m$ different suffixes must vanish, since in such cases it is impossible to avoid at least one g_{ij} with two free suffixes, and $\Sigma g_{ij} = 0$. For the same reason the only non-vanishing terms with $3m$ different suffixes are those containing expressions like

$$\Sigma g_{ij} g_{ik} g_{lu} g_{lv} g_{pq} g_{pr} \ldots = (\Sigma g_{ij} g_{ik})^m,$$

and terms with fewer different suffixes are of correspondingly lower order in N.

With $3m$ suffixes assigned there are $\binom{N-3m}{n-3m}$ ways of selecting the remaining $n-3m$ members of the sample, and the suffixes can be tied in $\frac{(2m)!\,2^{2m}}{m!\,(2!)^m}$ ways to give the same result. Dividing out by $\binom{N}{n}$ and noting that $\binom{N-3m}{n-3m}\Big/\binom{N}{n} \sim n^{3m}/N^{3m}$ when both N and n are large, the contribution of such terms to μ_{2m}, the $2m$th moment of $c^{(n)}$ about its mean, is found to be

$$\frac{n^{3m}}{N^{3m}} \frac{(2m)!}{m!} 2^m (\Sigma g_{ij} g_{ik})^m,$$

which is of order n^{3m}. Moreover, by the same argument, terms with $f < 3m$ different suffixes add contributions of order n^f which may be neglected.

Hence
$$\mu_{2m} \sim \frac{n^{3m}}{N^{3m}} \frac{(2m)!}{m!} 2^m (\Sigma g_{ij} g_{ik})^m,$$

the neglected terms being relatively $O(n^{-1})$.

13. For the odd moments let $r = 2m+1$. Similar considerations show that the non-vanishing terms of $\underset{n}{\Sigma}$ cannot have more than $3m+1$ different suffixes, and μ_{2m+1} is therefore of order n^{3m+1}.

Then since $c^{(n)}/n^{\frac{3}{2}}$ has even moments of unit order and odd moments of order $n^{-\frac{1}{2}}$, the odd moments may be neglected to that order. We conclude that $c^{(n)}$ is distributed normally for large n with variance

$$\frac{4n^3}{N^3} \Sigma g_{ij} g_{ik} = 4n^3 \operatorname{var} \tau_i,$$

and t is similarly normal with variance $(4/n) \operatorname{var} \tau_i$.

14. The fact that terms of order $n^{-\frac{1}{2}}$ have to be neglected suggests that the normal approximation only holds good for fairly large samples. This is not surprising since one would expect skewness to be an important property of the distribution of t when τ is not zero, if only for the reason that $|\,t\,|$ can never exceed unity. It seems worth while to examine the odd moments in more detail.

The dominant term of the $(2m+1)$th moment has $3m+1$ different suffixes, which can occur as
$$\Sigma g_{ij} g_{ik} g_{il} (\Sigma g_{uv} g_{uw})^{m-1} \quad \text{or} \quad \Sigma g_{ij} g_{ik} g_{jl} (\Sigma g_{uv} g_{uw})^{m-1}.$$
Both can be obtained in
$$\frac{(2m+1)!\,2^{2(m-1)}\,2^3}{(2!)^{m-1}\,3!(m-1)!} = \frac{(2m+1)!\,2^{m+2}}{3!(m-1)!}$$
distinct ways, and there are $\binom{N-3m-1}{n-3m-1}$ ways of selecting the sample with $3m+1$ suffixes assigned. The $(2m+1)$th moment of $c^{(n)}$ about its mean is therefore
$$\mu_{2m+1} \sim \frac{n^{3m+1}}{N^{3m+1}} \frac{(2m+1)!\,2^{m+2}}{3!(m-1)!} [\Sigma g_{ij} g_{ik} g_{il} + \Sigma g_{ij} g_{ik} g_{jl}] (\Sigma g_{ij} g_{ik})^{m-1},$$
ignoring terms of relative order $O(n^{-1})$. The corresponding moment of t is obtained to the same order on dividing by n^{4m+2}; it depends only on $\operatorname{var} t$ and $\mu_3(t)$, where
$$\mu_3(t) \sim \frac{8}{n^2 N^4} [\Sigma g_{ij} g_{ik} g_{il} + \Sigma g_{ij} g_{ik} g_{jl}] = \frac{4}{n^2} \frac{\Sigma g_{ij}(g_i + g_j)^2}{N^4},$$
where $g_i = \sum_{j=1}^{N} g_{ij}$. The distribution of t is thus specified to $O(n^{-1})$ by its first three moments.
The moment-generating function of the distribution of t in standard measure is
$$M(z) = \left(1 + \frac{\gamma_1}{3!} z^3\right) e^{-\frac{1}{2}z^2} [1 + O(n^{-1})],$$
where
$$\gamma_1 = \mu_3(t)/(\operatorname{var} t)^{\frac{3}{2}} = O(n^{-\frac{1}{2}}),$$
and the frequency distribution of $x = (t-\tau)/\sqrt{(\operatorname{var} t)}$ is*
$$f(x) = \left(1 - \frac{\gamma_1}{3!} \frac{d^3}{dx^3}\right) \frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}} [1 + O(n^{-1})]. \tag{14·1}$$

15. The effect of the γ_1 term in modifying the confidence limits based on normal theory can be seen in the following way. Let ξ be the normal deviate whose chance of being exceeded is $P(\xi)$. The chance of x exceeding ξ is, from (14·1),
$$F(\xi) = P(\xi) + \frac{\gamma_1}{6} (\xi^2 - 1) \frac{e^{-\frac{1}{2}\xi^2}}{\sqrt{(2\pi)}}.$$
If X is the correct limit such that $F(X) = P(\xi)$, it is readily proved by successive approximation that the formula
$$X = \xi + \frac{\gamma_1}{6} (\xi^2 - 1) \tag{15·1}$$
gives the appropriate value of X to $O(n^{-1})$. For example, the 5 and 1 % limits are respectively $\pm 1{\cdot}96 + 0{\cdot}474\gamma_1$ and $\pm 2{\cdot}58 + 0{\cdot}941\gamma_1$.

16. In practice the value of $\operatorname{var} t$ has to be estimated from the sample, and although its standard error can be shown to be $O(n^{-\frac{1}{2}})$ by the kind of argument already used, it is not known how large the sample has to be before the error in estimating the variance can be safely ignored. It is best to use the unbiased formula
$$\operatorname{var} t = \frac{1}{n(n-1)(n-2)(n-3)} \left\{4\Sigma c_i^2 - \frac{2(2n-3)}{n(n-1)} c^2 - 2n(n-1)\right\} \tag{16·1}$$
(which is easily proved) in calculating $\operatorname{var} t$ from the sample, especially if the standard error of the mean value of t from a number of small samples is required.

* Note that the approximation error in $f(x)$ is relatively $O(n^{-1})$, a stronger result than would be obtained from a Gram-Charlier approximation based on the first three moments only.

As the term in γ_1 is a small correction it is perhaps sufficient in moderate samples to take

$$G = \tfrac{1}{2}\Sigma g_{ij}(g_i+g_j)^2 = \tfrac{1}{2}\Sigma c_{ij}(c_i+c_j)^2 - \frac{5c\Sigma c_i^2}{n} + \frac{3c^2}{n^3}, \tag{16·2}$$

and

$$\mu_3(t) = \frac{8}{n^6}G, \quad \gamma_1 = \mu_3(t)/(\mathrm{var}\, t)^{\frac{3}{2}}, \tag{16·3}$$

where the first term in G is the sum of $c_{ij}(c_i+c_j)^2$ over all values of $i>j$. The unbiased formula for $\mu_3(t)$ involves some rather tedious computation.

17. To illustrate the methods of the paper we consider an actual example.

A set of thirty wool samples were visually graded in order of fibre fineness by three assessors. The mean fibre diameter for each wool sample was also determined by direct measurement. Table 1 shows the measured order (M) compared with that of the three assessors (A, B, C), in ascending order of experience.

Table 1

M	A	B	C	M	A	B	C
1	5	2	1	16	12	14	16
2	4	5	2	17	10	18	15
3	9	6	6	18	30	21	25
4	3	1	3	19	22	26	24
5	6	7	4	20	16	22	19
6	2	4	5	21	21	16	18
7	15	19	10	22	29	20	23
8	18	3	12	23	28	25	22
9	8	8	7	24	19	27	26
10	11	9	8	25	23	28	21
11	17	13	9	26	20	23	27
12	13	10	11	27	7	24	20
13	24	17	17	28	26	29	28
14	14	12	14	29	27	15	30
15	1	11	13	30	25	30	29

The method of working will be seen from the c_{ij} matrix for the MA correlation shown in Table 2.

The correlations of the assessors' orders with the measured order are found to be

$$t_A = 0{\cdot}490, \quad t_B = 0{\cdot}724, \quad t_C = 0{\cdot}816.$$

(i) Consider first the maximum confidence limits given by (10·2). The 5 % limits are

$$-0{\cdot}02 < t_A < 0{\cdot}80, \quad 0{\cdot}23 < t_B < 0{\cdot}92, \quad 0{\cdot}34 < t_C < 0{\cdot}96.$$

Again, using the transformation $w = \sin^{-1} t$, the 5 % limits are

$$0{\cdot}01 < t_A < 0{\cdot}85, \quad 0{\cdot}30 < t_B < 0{\cdot}97, \quad 0{\cdot}45 < t_C < 0{\cdot}99.$$

The values of w are $\quad w_A = 0{\cdot}512, \quad w_B = 0{\cdot}810, \quad w_C = 0{\cdot}954.$

The greatest difference is 0·442, and the upper limit to its standard error is $\sqrt{(4/n)} = 0{\cdot}365$, so on these grounds the difference between A and C would not be judged significant.

The 5 % limits are very wide, and the lack of significance is disappointing since C was known to be an expert appraiser while A is relatively inexperienced, and one would have expected an obvious difference between them.

(ii) The variances estimated from the unbiased formula (16·1) are

$$\operatorname{var} t_A = 0{\cdot}006630, \quad \operatorname{var} t_B = 0{\cdot}005067, \quad \operatorname{var} t_C = 0{\cdot}002198.$$

The estimated standard errors are therefore

$$s_A = 0{\cdot}081, \quad s_B = 0{\cdot}071, \quad s_C = 0{\cdot}047.$$

The 5 % confidence limits, assuming normality, are

$$0{\cdot}33 < t_A < 0{\cdot}65, \quad 0{\cdot}58 < t_B < 0{\cdot}86, \quad 0{\cdot}72 < t_C < 0{\cdot}91.$$

Table 2

														c_{ij}																c_i
0	−	+	−	+	−	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	21
−	0	+	−	+	−	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	21
+	+	0	−	−	−	+	+	−	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	17
−	−	−	0	+	−	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	19
+	+	−	+	0	−	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	23
−	−	−	−	−	0	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	17
+	+	+	+	+	+	0	+	−	−	+	−	+	−	−	−	−	+	+	+	+	+	+	+	+	+	−	+	+	+	13
+	+	+	+	+	+	+	0	−	−	−	−	+	−	−	−	−	+	+	−	+	+	+	+	+	+	−	+	+	+	9
+	+	−	+	+	+	−	−	0	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	19
+	+	+	+	+	+	−	−	+	0	+	+	+	+	−	+	−	+	+	+	+	+	+	+	+	+	−	+	+	+	19
+	+	+	+	+	+	+	−	+	+	0	−	+	−	−	−	−	+	+	−	+	+	+	+	+	+	−	+	+	+	13
+	+	+	+	+	+	−	−	+	+	−	0	+	+	−	−	−	+	+	+	+	+	+	+	+	+	−	+	+	+	15
+	+	+	+	+	+	+	+	+	+	+	+	0	−	−	−	−	+	−	−	−	+	+	−	−	−	−	+	+	+	7
+	+	+	+	+	+	−	−	+	+	−	+	−	0	−	−	−	+	+	+	+	+	+	+	+	+	−	+	+	+	13
−	−	−	−	−	−	−	−	−	−	−	−	−	−	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	1
+	+	+	+	+	+	−	−	+	+	−	−	−	−	+	0	−	+	+	+	+	+	+	+	+	+	−	+	+	+	13
+	+	+	+	+	+	−	−	+	−	−	−	−	−	+	−	0	+	+	+	+	+	+	+	+	+	−	+	+	+	11
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	−	−	−	−	−	−	−	−	−	−	−	−	5
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	0	−	−	+	+	−	+	−	−	+	+	+	15
+	+	+	+	+	+	+	−	+	+	−	+	−	+	+	+	+	−	−	0	+	+	+	+	+	+	−	+	+	+	17
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	−	+	0	+	+	−	+	+	−	+	+	+	19
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	0	−	−	−	−	−	−	−	−	11
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	0	−	−	−	−	−	−	−	11
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	−	+	−	−	−	0	+	+	−	+	+	+	15
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	+	+	+	−	−	+	0	−	−	+	+	+	17
+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	+	−	−	+	+	−	−	+	−	0	−	+	+	+	15
+	+	−	+	+	+	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	0	+	+	+	−11
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	−	+	+	+	+	0	+	−	21
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	−	+	+	+	+	+	0	−	21
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	−	−	+	+	+	+	−	−	0	19

$c = 426$
$n = 30$
$\Sigma c_i^2 = 7470$

Moreover, we should judge A and C to be significantly different at the 1 % level, and A and B at the 5 % level. How far these conclusions are valid depends, of course, on the accuracy of the variance estimates, but the conclusions seem to agree with what might have been expected from prior knowledge of the assessors' capabilities.

(iii) The values of γ_1 calculated from (16·2) and (16·3) are

$$\gamma_1(A) = -0{\cdot}32, \quad \gamma_1(B) = -0{\cdot}35, \quad \gamma_1(C) = -0{\cdot}38.$$

The distributions would not appear to be very skew, and the distribution of the difference of two t's is probably nearly normal. The adjusted 5 % limits are, from (15·1),

$$0{\cdot}32 < t_A < 0{\cdot}64, \quad 0{\cdot}57 < t_B < 0{\cdot}85, \quad 0{\cdot}72 < t_C < 0{\cdot}90.$$

APPENDIX

1. The question arises whether a particular parental form exists for which the variance of t assumes the upper limit $2(1-\tau^2)/n$. We surmise, though we cannot prove, that the maximum possible variance is attained when the parent ranking has a 'canonical' form obtained in the following way. Consider again the ranking

$$5 \quad 2 \quad 3 \quad 1 \quad 6 \quad 7 \quad 8 \quad 9 \quad 4.$$

The number of positive pairs S is 26, so that $t = 0{\cdot}44$. Let us transform this so as to bring the 1 to the beginning of the ranking but move the 9 so as to preserve the number S at 26. The 1 passes over three members to go to the beginning and hence adds 3 to the score. The 9 must, therefore, proceed to the left over three numbers so as to subtract 3 from the score and we reach

$$1 \quad 5 \quad 2 \quad 3 \quad 9 \quad 6 \quad 7 \quad 8 \quad 4.$$

Now operate similarly with 2 and 9, reaching

$$1 \quad 2 \quad 5 \quad 9 \quad 3 \quad 6 \quad 7 \quad 8 \quad 4.$$

Had our 9 been contiguous to the 1 and incapable of moving farther to the left we should have moved the 8 and so on. Proceeding with the process by moving back the 3 and the 9 and 8 we reach

$$1 \quad 2 \quad 3 \quad 9 \quad 5 \quad 6 \quad 8 \quad 7 \quad 4,$$

and again

$$1 \quad 2 \quad 3 \quad 4 \quad 9 \quad 8 \quad 7 \quad 6 \quad 5.$$

All the lower numbers 1 to 4 are in the right order and the remainder are in the inverse order. We call this ranking the 'canonical' order for given S (or t). It is not always possible to reduce a given ranking to canonical order, but there cannot be more than one individual out of place.

2. Consider the effect of a series of transformations leading to the canonical form. The first process, that of moving 1 and 9, will increase the value of S for some samples involving 1 but not 9 (leaving the others unchanged), will decrease the value of S for some samples involving 9 but not 1 (leaving the others unchanged), and will, in general, not alter those involving both 1 and 9. Similarly for 2 and 8, and so on. The effect of the transformation is thus to increase the values of S containing the lower numbers 1, 2, 3, etc., and to decrease those containing 9, 8, 7, etc. These values of S are themselves, in the canonical form, the greatest or least as the case may be. Consequently the progress to the canonical form is accompanied by increases in the number of high values of S and increases in the number of lower values, and one might expect the spread of the distribution to tend to a maximum. In the example quoted, the distributions of S in samples of 3 for the successive rankings are:

Values of S	Frequencies f				
0	2	3	3	6	10
1	15	13	16	10	—
2	34	35	29	32	40
3	33	33	36	36	34
Totals	84	84	84	84	84

The sums ΣfS are all equal to 182. The sums ΣfS^2 are respectively 448, 450, 456, 462 and 466, showing the canonical ranking to have the largest variance of the five.

3. There is, however, another way of carrying out this process. If the parent ranking is inverted, τ becomes $-\tau$, but the variance of samples of n drawn from the inverted ranking remains the same, by symmetry. We may then reduce the inverted ranking to its canonical form and reinvert it so that its coefficient is again τ. This ranking we call the inverse canonical form. It will be shown that for large N, when $\tau > 0$ the inverse canonical form yields a larger variance for t than the direct canonical form.

Even in the example already quoted, the inverse canonical ranking (with one member out of place) is

$$3 \quad 4 \quad 2 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 1,$$

which has a distribution

Values of S	f
0	2
1	27
2	10
3	45
Total	84

The sum $\Sigma f S^2$ is now 472, which is greater than the previous maximum 466.

4. Consider the canonical case when there are N members altogether, R at the beginning in the right order, and $N-R$ in the inverse order. If we select $n-j$ members from the R and j from the $N-R$ the value of S for the sample of n is $\frac{1}{2}n(n-1)-\frac{1}{2}j(j-1)$, and the relative frequency of $U = \frac{1}{2}n(n-1)-S$ is $\binom{R}{n-j}\binom{N-R}{j}\Big/\binom{N}{n}$. Now suppose that N tends to infinity and R/N to the ratio p. The relative frequency of $U = \frac{1}{2}j(j-1)$ tends in the limit to

$$\binom{n}{j} p^{n-j} q^j,$$

where $q = 1-p$. The mean value of U is then

$$\sum_0^n \tfrac{1}{2}j(j-1)\binom{n}{j} p^{n-j}q^j = \tfrac{1}{2}n(n-1)\,q^2,$$

and since
$$t = 1-\frac{2U}{\frac{1}{2}n(n-1)},$$

we must have
$$q = \{\tfrac{1}{2}(1-\tau)\}^{\frac{1}{2}}. \tag{4·1 A}$$

The variance of U is
$$\operatorname{var} U = n(n-1)\,pq^2\{nq-\tfrac{1}{2}(1-3q)\}, \tag{4·2 A}$$

and so
$$\operatorname{var} t \doteq 16pq^2\{nq-\tfrac{1}{2}(1-3q)\}/n(n-1). \tag{4·3 A}$$

5. If now the inverted parent ranking is reduced to canonical form, giving ratios p', q' corresponding to p and q, we shall have

$$q' = \sqrt{[\tfrac{1}{2}(1+\tau)]} \tag{5·1 A}$$

and
$$\operatorname{var} t' = 16p'q'^2\{nq'-\tfrac{1}{2}(1-3q')\}/n(n-1). \tag{5·2 A}$$

Then since $q^2+q'^2 = 1$,

$$\operatorname{var} t' - \operatorname{var} t = \frac{16(n+2)}{n(n-1)}(q'-q)(1-q)(1-q'). \tag{5·3 A}$$

When τ is positive, $q' > q$ and $\operatorname{var} t'$ exceeds $\operatorname{var} t$.

This result suggests that the maximum variance may be attained by the inverse canonical ranking when $\tau > 0$ and by the direct canonical ranking when $\tau < 0$. With this choice of parent ranking the variance of t for large n is

$$\operatorname{var} t \sim \frac{4\sqrt{2}}{n}(1+|\tau|)^{\frac{1}{2}}\{1-\sqrt{[\tfrac{1}{2}(1+|\tau|)]}\}. \tag{5·4A}$$

It is interesting to compare (5·4A) with our upper limit of $2(1-\tau^2)/n$. Their ratio is $\{2(1+|\tau|)\}^{\frac{1}{2}}/[1+\{\frac{1}{2}(1+|\tau|)\}^{\frac{1}{2}}]$, which varies from $2(\sqrt{2}-1) = 0·83$ when $\tau = 0$ to 1 when $\tau = 1$. Evidently the upper limit to the variance cannot be much improved, since an actual ranking has been found whose variance approximates to it for all values of τ, when n is not too small.

REFERENCES

DANIELS, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, 33, 129.

KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, 1. 2nd ed. 1945. London: Charles Griffin and Co.

TESTING FOR NORMALITY

By R. C. GEARY, *Cambridge University Department of Applied Economics*

1. Introduction

The present communication, one of a series, has two main objectives:

(1) To show that probabilities derived from the well-known analyses of variance and other 'small sample' tables, which postulate universal normality, may differ seriously from the true probabilities when the universes are non-normal, even, in some cases, when the degree of non-normality is not considerable.

(2) To determine the most efficient tests of normality from a wide field of alternative symmetrical tests.

It may be useful to summarize very briefly previous work in so far as it is strictly relevant to this study.* The modern theory may be regarded as having been initiated by Karl Pearson who, in 1895, found the first approximation (i.e. to $n^{-\frac{1}{2}}$) to the variances and covariance of $\sqrt{b_1}$ and b_2 for samples drawn at random from any universe and, assuming that the $\sqrt{b_1}$ and b_2 were distributed jointly with normal probability, constructed 'probability ellipses' from which the probability of the same values occurring; had the universe, in fact, been normal, could be inferred very approximately. A considerable advance in moment determination was made by C. C. Craig (1928). In 1929, R. A. Fisher, in inventing cumulants, simple functions of the sample moments, and formulating rules for finding their semi-invariants, developed incidentally a technique for expanding to several terms in $1/n$ the moments of $\sqrt{b_1}$ and b_2 when the universe was normal. This paper was followed soon after by another (1930), fundamental for all succeeding work on this subject, in which R. A. Fisher ingeniously applied combinatorial technique to the finding of exact values of the moments of normal $\sqrt{b_1}$ and b_2, and gave *inter alia* the values of the second, fourth and sixth moments of $\sqrt{b_1}$ and of the first three moments of b_2. The fourth semi-invariant, together with many other normal semi-invariants of b_2, was determined by J. Wishart in 1930, and a further advance in R. A. Fisher's technique was made jointly by R. A. Fisher & J. Wishart in 1930. In 1932 Joseph Pepper gave the eighth normal moment of $\sqrt{b_1}$. Using R. A. Fisher's rules C. T. Hsu and D. N. Lawley in 1940 gave the exact values for normal random samples of the fifth and sixth moments of b_2. Using a method due to R. C. Geary (1933) (applying C. C. Craig's ideas (1928) to the normal problem), R. C. Geary & J. P. G. Worlledge have recently (1946) found the seventh moment of b_2.

So much for moment determination. In 1930, E. S. Pearson used appropriate Pearson-type curves, applied to R. A. Fisher's (1929) approximations of the semi-invariants, to find approximate frequency distributions of $\sqrt{b_1}$ and b_2. From the frequency distributions he computed a table of 1 % and 5 % probability points at intervals for n from 50 to 5000 for $\sqrt{b_1}$ and for n from 100 to 5000 for b_2.

Since at the time the prospect seemed remote of determining the frequency of normal b_2 on which reliance could be reposed for samples of moderate sizes, R. C. Geary (1935)† suggested that the ratio, a, of mean deviation to standard deviation computed from the origin

* An excellent account of the development of moment theory up to the year 1930 was given by J. Wishart (1930).

† The author was informed by M. Fréchet that this test was suggested by Bertrand, but has been unable to check the reference.

might be used as a test of normality, and gave the 1 and 5 % probability points for this test at intervals for normal samples of 6–100. E. S. Pearson compared experimentally Geary's test with b_2 and suggested, for samples so large that comparison could safely be made, that b_2 was probably somewhat more sensitive than a, a suggestion which will be examined theoretically in this communication. In 1935 also, R. C. Geary showed that there was a high (negative) correlation for normal samples between $a(1)$ (see 3·1) and b_2 for normal samples, and argued therefrom that the former should be nearly as efficient as b_2. In 1936, R. C. Geary gave a table of 1, 5 and 10 % probability points of $a(1)$ at intervals for samples of 11–1001. In 1938, a brochure by R. C. Geary & E. S. Pearson was published by the Biometrika Office entitled *Tests of Normality*, giving tables and diagrams of probability points of $a(1)$ $\sqrt{b_1}$ and b_2. There is considerable literature dealing with the effect of universal non-normality on the normal tests, mostly by way of particular numerical examples: a selection of papers on this subject is included in the list of references at the end of the paper.

2. Effect of non-normality

(a) *The z-test*

The effect of universal non-normality will first be considered in relation to the z-test. If $x_1, x_2, \ldots, x_{n'}$ and $y_1, y_2, \ldots, y_{n''}$ are two independent samples drawn at random from the same universe (normal or non-normal) it is easy to show that, if

$$z = \tfrac{1}{2}\log\frac{n''-1}{n'-1}\frac{\sum\limits_{i=1}^{n'}(x_i-\bar{x})^2}{\sum\limits_{i=1}^{n''}(y_i-\bar{y})^2} = \tfrac{1}{2}\log\frac{s'^2}{s''^2}, \tag{2·1}$$

then

$$\sigma_z^2 = \frac{(\beta_2-1)}{4}\left(\frac{1}{n'}+\frac{1}{n''}\right) = M_2, \tag{2·2}$$

when both n' and n'' are so large that terms in n' and n'' of degree less than -1 are regarded as negligible. This is an obvious generalization of the approximate formula given by R. A. Fisher* for normal samples, namely,

$$\sigma_z^2 = \frac{1}{2}\left(\frac{1}{n'}+\frac{1}{n''}\right) = M_2^0. \tag{2·3}$$

It may be useful also to give formulae for the first and second moments from zero for z when the two random samples are drawn not necessarily from the same universes, though both universes have mean zero and the same variance λ_2:

$$\left.\begin{aligned}
2M_1' = {} & -\frac{1}{2\lambda_2^2}\left(\frac{\lambda_4'}{n'}+\frac{2\lambda_2^2}{n'-1}\right)+\frac{1}{2\lambda_2^2}\left(\frac{\lambda_4''}{n''}+\frac{2\lambda_2^2}{n''-1}\right)+\frac{1}{3\lambda_2^3}\left[\left(\frac{\lambda_6'}{n'^2}-\frac{\lambda_6''}{n''^2}\right)+12\lambda_2\left(\frac{\lambda_4'}{n'^2}-\frac{\lambda_4''}{n''^2}\right)\right.\\
& \left.+4\left(\frac{\lambda_3'^2}{n'^2}-\frac{\lambda_3''^2}{n''^2}\right)+8\lambda_2^3\left(\frac{1}{n'^2}-\frac{1}{n''^2}\right)\right]-\frac{3}{4\lambda_2^4}\left[\frac{(\lambda_4'+2\lambda_2^2)^2}{n'^2}-\frac{(\lambda_4''+2\lambda_2^2)^2}{n''^2}\right]+\ldots,\\
4M_2' = {} & \frac{1}{\lambda_2^2}\left[\left(\frac{\lambda_4'}{n'}+\frac{\lambda_4''}{n''}\right)+2\lambda_2^2\left(\frac{1}{n'-1}+\frac{1}{n''-1}\right)\right]\\
& -\frac{1}{\lambda_2^3}\left[\left(\frac{\lambda_6'}{n'^2}+\frac{\lambda_6''}{n''^2}\right)+12\lambda_2\left(\frac{\lambda_4'}{n'^2}+\frac{\lambda_4''}{n''^2}\right)+4\left(\frac{\lambda_3'^2}{n'^2}+\frac{\lambda_3''^2}{n''^2}\right)+8\lambda_2^3\left(\frac{1}{n'^2}+\frac{1}{n''^2}\right)\right]\\
& +\frac{1}{12\lambda_2^4}\left[\frac{33}{n'^2}(\lambda_4'+2\lambda_2^2)^2+\frac{33}{n''^2}(\lambda_4''+2\lambda_2^2)^2-\frac{6}{n'n''}(\lambda_4'+2\lambda_2^2)(\lambda_4''+2\lambda_2^2)\right]+\ldots,
\end{aligned}\right\} \tag{2·4}$$

* *Statistical Methods for Research Workers*, 8th ed. p. 219.

where the λ's indicate semi-invariants of the two universes of the orders indicated. In these formulae, in effect, terms to order -2 in n', n'' are retained.

When both samples are large the frequency distribution of z will approach normality provided that μ_4 is finite. The effect of universal kurtosis can accordingly be assessed in a very rudimentary manner from (2·2) and (2·3). The z-deviate ζ corresponding to, say, the $2\frac{1}{2}$ % normal probability point is

$$\zeta = 1{\cdot}9600\sqrt{M_2^0}. \tag{2·5}$$

If, however, the universe were not normal and had, *in fact*, a variance M_2 with $\beta_2 \neq 3$, the *actual* probability of a deviation in excess of ζ in absolute value would be, not 0·05, but the normal probability appropriate to a unit variance deviate of $\zeta M_2^{-\frac{1}{2}}$. On this consideration the actual probabilities for different values of β_2, where the assumed probability is 0·05, are shown in the fifth column of Table 1.

Table 1. *Effect on probability of z of change in universal kurtosis, for large samples*

β_2	M_2^0/M_2	$\sqrt{(M_2^0/M_2)}$	$1{\cdot}9600\sqrt{(M_2^0/M_2)}$	Actual probability
1·5	4	2	3·9200	0·000089
2	2	1·4142	2·7718	0·0056
2·5	1·3333	1·1547	2·2632	0·024
3	**1**	**1**	**1·9600**	**0·050**
3·5	0·8000	0·8944	1·7530	0·080
4	0·6667	0·8165	1·6003	0·110
4·5	0·5714	0·7559	1·4816	0·138
5	0·5000	0·7071	1·3859	0·166
5·5	0·4444	0·6667	1·3065	0·191
6	0·4000	0·6325	1·2397	0·215

The table shows that, if the universe from which the samples are drawn has $\beta_2 = 6$, the true probability is about 1 in 5 instead of the assumed 1 in 20. It is, of course, true that universes with so large a kurtosis are unusual. This view cannot be held of the range 2·5–4 for β_2 in which the probability, assumed to be 0·05, can be anything, in fact, from 0·024 to 0·110. Accordingly, if universal kurtosis is markedly negative, use of the standard table masks significant differences; if kurtosis is positive the standard table exaggerates these differences. Unless systematic tests have established that kurtosis is negligible the standard table should not be used for testing significant differences in variance.

The foregoing analysis gives a theoretical explanation of the striking experimental results of E. S. Pearson (1931*b*) working, however, with a test function

$$x = \sum_{i=1}^{n'} (x_i - \bar{x})^2 \Bigg/ \left\{ \sum_{i=1}^{n'} (x_i - \bar{x})^2 + \sum_{i=1}^{n''} (y_i - \bar{y})^2 \right\}$$

and with sample sizes $n' = 5$ and $n'' = 20$, smaller than those contemplated in the present analysis. With 500 samples Pearson showed that when the frequency at the two tails together expected from normal theory was 15·4 (=probability 0·0308) the frequencies actually found in symmetrical universes with $\beta_2 = 2{\cdot}5$, 4·1 and 7·1 respectively were 7, 39 and 47, equivalent to probabilities of 0·014, 0·078 and 0·094.

If tests of normality indicate universal kurtosis, either of two courses might be adopted:

(i) Assume that z is normally distributed with variance M_2 computed from (2·2) with (β_2-3) estimated as k_4/k_2^2 from the sample, k_2 and k_4 being R. A. Fisher's (1929) cumulant functions.

(ii) Enter the standard table, not with z computed from the samples but with $z\sqrt{(M_2^0/M_2)}$, estimating M_2 as in (i).

Both of these procedures are, of course, open to the objection that, unless the samples are extremely large the estimate of β_2 is unlikely to be accurate; the real β_2 might be larger or smaller than the estimate. Any probabilistic inferences should accordingly be accepted with reserve.

It is fortunate that the condition specified in the foregoing paragraphs, namely, that the numbers in the two samples are both large, rarely applies in practical applications. It more usually happens that the number of classes is small, whereas the number per class is relatively large. In this case E. S. Pearson (1931 *b*) has shown the first approximation to σ_z^2 is independent of β_2, from which he inferred that the actual probability when the total number of samples was large was inconsiderably influenced by kurtosis. In view of the foregoing analysis it seemed to the writer desirable to carry the inquiry a stage further.

Suppose, then, that k samples are drawn at random from the same universe, n_j in the jth sample, the total $\sum_j n_j = n$. It is assumed that n is so large that terms in n^{-2} are negligible, that the number of samples k is small, and that all the n_j are of the same order of magnitude as n, i.e. that if

$$n_j = \pi_j n, \quad \sum_{j=1}^{k} \pi_j = 1, \tag{2·6}$$

none of the π_j is negligibly small.

Using R. A. Fisher's cumulant notation with subscript to indicate the sample from which the cumulants were computed, the mean for the jth sample is written k_{1j} and its variance k_{2j}. Then

$$z = \tfrac{1}{2}\log\frac{X}{Y}, \tag{2·7}$$

where
$$(k-1)\,X = \sum_j n_j(k_{1j}-k_1)^2 = \sum n_j k_{1j}^2 - \frac{1}{n}\Sigma^2 n_j k_{1j},$$

so that
$$\frac{k-1}{n}X = \sum \pi_j(1-\pi_j)\,k_{1j}^2 - 2\sum_{j<j'}\sum \pi_j\pi_{j'}k_{1j}k_{1j'},$$

and
$$(n-k)\,Y = \sum_j (n_j-1)\,k_{2j},$$

so that
$$Y = \Sigma\phi_j k_{2j},$$

where
$$\phi_j = \frac{n_j-1}{n-k}.$$

Without loss of generality let the universal mean be zero and the variance unity. It may easily be shown that

$$EX = EY = 1.$$

Set
$$w = \frac{X}{Y} = \frac{\bar{X}+(X-\bar{X})}{\bar{Y}+(Y-\bar{Y})} = \{1+(X-1)\}\{1+(Y-1)\}^{-1}.$$

Then
$$\left.\begin{aligned} w &= \{1+(X-1)\}\{1-(Y-1)+(Y-1)^2-(Y-1)^3+\ldots\},\\ w^2 &= \{1+(X-1)\}^2\{1-2(Y-1)+3(Y-1)^2-4(Y-1)^3+\ldots\}.\end{aligned}\right\} \tag{2·8}$$

We shall compute the approximate values of Ew and Ew^2, i.e. the values to order n^{-1}; the symbol $\simeq$ denotes 'equal to, to approximation required'. From values of the variances and covariances given by E. S. Pearson (1931*b*) in his equations (9)–(11), we have

$$\left.\begin{aligned} E(X-1)^2 &= \frac{2}{k-1}+(1-2k+\alpha_{-1}-1)\frac{\lambda_4}{n(k-1)^2},\\ E(X-1)(Y-1) &\simeq \frac{\lambda_4}{n},\\ E(Y-1)^2 &\simeq \frac{\lambda_4+2}{n},\end{aligned}\right\} \tag{2·9}$$

with
$$\alpha_c = \sum_j \pi_j^c.$$

We require $\left(\frac{k-1}{n}\right)^2 X^2 \simeq \sum_j \pi_j^2(1-\pi_j)^2 k_{1j}^4 - 4\sum_j\sum_{j'} \pi_j^2(1-\pi_j)\pi_{j'}k_{1j}^3k_{1j'}$

$$+2\sum_j\sum_{j'}\pi_j\pi_{j'}(1-\pi_j-\pi_{j'}+3\pi_j\pi_{j'})k_{1j}^2k_{2j'}^2 - 4\sum_j\sum_{j'}\sum_{j''}\pi_j\pi_{j'}\pi_{j''}(1-3\pi_j)k_{1j}^2k_{1j'}k_{1j''},$$

$$Y-1 = \Sigma\phi_j(k_{2j}-1) = \Sigma\phi_j k'_{2j}, \quad \text{say},$$

remembering that, by definition of cumulants,

$$Ek_{2j} = \lambda_2 = 1.$$

Also
$$(Y-1)^2 = \Sigma\phi_j^2 k_{2j}'^2 + 2\sum_{j>j'}\sum \phi_j\phi_{j'}k'_{2j}k'_{2j'}.$$

It will be useful for what follows to note that

$$\phi_j \simeq \pi_j.$$

Using R. A. Fisher's formulae (1929) for formation of joint semi-invariants of k_1 and k_2, and noting that the k samples are independent, we find from the foregoing

$$\left.\begin{aligned} n(k-1)\,EX(Y-1)^2 &\simeq (k-1)(\lambda_4+2),\\ n(k-1)^2\,EX^2(Y-1) &\simeq 2(k^2-1)\lambda_4,\\ n(k-1)^2\,EX^2(Y-1)^2 &\simeq (k^2-1)(\lambda_4+2).\end{aligned}\right\} \tag{2·10}$$

Then, from (2·8), (2·9), (2·10),

$$\left.\begin{aligned} Ew &\simeq 1+\frac{2}{n},\\ Ew^2 &\simeq \frac{k+1}{k-1}+\frac{1}{n(k-1)^2}\{6(k^2-1)-(k^2+2k-2-\alpha_{-1})\lambda_4\}.\end{aligned}\right\} \tag{2·11}$$

These are the formulae required. It will be noted

(i) that the terms free of n^{-1} are independent of λ_4, which is equivalent to E. S. Pearson's result (1931*b*);

(ii) that the formulae (2·11) agree with the normal values

$$\left.\begin{aligned} E_0 w &= \left(1-\frac{2}{n-k}\right)^{-1} \simeq 1+\frac{2}{n},\\ E_0 w^2 &= \frac{k+1}{k-1}\left(1-\frac{2}{n-k}\right)^{-1}\left(1-\frac{4}{n-k}\right)^{-1} \simeq \frac{k+1}{k-1}\left(1+\frac{6}{n}\right),\end{aligned}\right\} \tag{2·12}$$

to n^{-1} when $\lambda_4 = 0$;

(iii) the approximations at (2·11) are free of λ_3.

The approximations at (2·11) tend to confirm E. S. Pearson's result that, when n is large compared with k, the effect of universal kurtosis is unimportant. It would be useful, however, to compute the approximate true probability for different values of k, n, λ_4 and α_{-1}. For this and for subsequent work the following lemma* will be found useful:

If $f(x)$ and $\phi(x)$ are two frequency densities with semi-invariants L_m and L'_m $(m = 1, 2, \ldots)$, respectively, then, formally,

$$f(x) = \exp\left\{\sum_{m=1}^{\infty} \frac{(L_m - L'_m)}{m!}\left(-\frac{d}{dx}\right)^m\right\}\phi(x). \tag{2·13}$$

For the present application take as generating function ϕ the frequency distribution of w in the normal case, i.e.

$$\phi(w) = \frac{\left(\frac{n-3}{2}\right)!\left(\frac{k-1}{n-k}\right)^{\frac{1}{2}(k-1)}}{\left(\frac{k-3}{2}\right)!\left(\frac{n-k-2}{2}\right)!} w^{\frac{1}{2}(k-3)}\left\{1+\frac{(k-1)\,w}{(n-k)}\right\}^{-\frac{1}{2}(n-1)}, \tag{2·14}$$

and, from (2·11),

$$L_2 - L'_2 \simeq -\frac{(k^2+2k-2-\alpha_{-1})\,\lambda_4}{n(k-1)^2}. \tag{2·15}$$

Assume that $\qquad L_m - L'_m \simeq 0 \quad (m \neq 2).$

Then if the 'normal theory' probability corresponding to the sample value w be p, the approximate 'true' probability, subject to (2·15), will be about $(p+p')$, where p' is given by

$$p' = \frac{(L_2 - L'_2)}{2}\int_w^{\infty} \phi''(w)\,dw = -\frac{(L_2 - L'_2)}{2}\phi'(w). \tag{2·16}$$

The term p', of course, merely corrects for the non-normal term in n^{-1} in the variance of z; it takes no account of corrections due to terms of higher (negative) orders in n or even of non-normal terms in n^{-1} in semi-invariants L_m $(m > 2)$. The calculation is designed merely to show whether the standard table probability requires correction for universal kurtosis; this will appear if p' is of the order of magnitude of p.

(b) *The t-test*

In Geary's 1936 paper the expansion to terms in n^{-2} of the first *four* moments of t, where

$$t = n^{\frac{1}{2}} k_1 / k_2^{\frac{1}{2}}, \tag{2·17}$$

were given. Following are the first *six* semi-invariants L of t to the same approximation as in the earlier paper:

$$\left.\begin{aligned}
L_1 &\simeq -\frac{1}{n^{\frac{1}{2}}}\left\{\frac{\lambda_3}{2}+\frac{3}{16n}(2\lambda_3 - 2\lambda_5 + 5\lambda_3\lambda_4)\right\}+\ldots,\\
L_2 &\simeq 1+\tfrac{1}{4}(8+7\lambda_3^2)\,n^{-1}+(6-2\lambda_4-\tfrac{3}{8}\lambda_3^2-\tfrac{45}{8}\lambda_3\lambda_5+\tfrac{177}{16}\lambda_3^2\lambda_4)\,n^{-2},\\
L_3 &\simeq -2\lambda_3 n^{-\frac{1}{2}}-(9\lambda_3-3\lambda_5+\tfrac{15}{4}\lambda_3\lambda_4+\tfrac{83}{8}\lambda_3^3)\,n^{-\frac{3}{2}},\\
{}^{\dagger}L_4 &\simeq (6-2\lambda_4+12\lambda_3^2)\,n^{-1}+(54-18\lambda_4+4\lambda_6+75\lambda_3^2-63\lambda_3\lambda_5-6\lambda_4^2+81\lambda_3^2\lambda_4+\tfrac{699}{8}\lambda_3^4)\,n^{-2},\\
L_5 &\simeq -(60\lambda_3-6\lambda_5-20\lambda_3\lambda_4+105\lambda_3^3)\,n^{-\frac{3}{2}},\\
L_6 &\simeq (240-120\lambda_4+577\tfrac{1}{2}\lambda_3^2+16\lambda_6-210\lambda_3\lambda_5-150\lambda_3^2\lambda_4+1200\lambda_3^4)\,n^{-2}.
\end{aligned}\right\} \tag{2·18}$$

Throughout this subsection we take $\qquad \lambda_m = \lambda'_m / \lambda_2'^{\frac{1}{2}m},$

* Due to Charlier and termed the "Differential Series" by the Scandinavian School.

† 1936 formula corrected.

where the λ'_m are the semi-invariants of the parent universe. For these expressions terms in $n^{-\frac{5}{2}}$ are neglected. They were derived from the moments (from zero) M'_i of t, which were obtained by the method described in the 1936 paper. It will be noted that, to the approximation used, the expressions involve only the first six semi-invariants of the parent universe. When the parent universe is normal all the λ_i $(i > 2)$ are zero. The magnitude of the numerical coefficients in the foregoing approximate expressions for the L_i indicate that, when the universal values of the λ_i, particularly those of uneven order, are not very small, the frequency distribution of t may differ appreciably from the classical Gosset-Fisher (1908, 1925) distribution.

The formal Gram-Charlier expression for the frequency of t could, of course, be written down at once from (2·18). It is doubtful, however, if the Gaussian can be regarded as the most appropriate generating function for the frequency of t because, even when the parent universe is normal, the semi-invariants T'_{2m} of the higher even orders are large for moderate values of n. For example,

$$L'_4/L'^2_2 = 6/(n-5), \quad L'_6/L'^3_2 = 240/(n-5)(n-7).$$

It is proposed to use (2·13) for finding the approximate frequency with

$$\phi(t) = T(t;\, n) = \left(\frac{n-2}{2}\right)!\left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n} \Big/ \left(\frac{n-3}{2}\right)!\,(\pi\overline{n-1})^{\frac{1}{2}}, \tag{2·19}$$

the Gosset-Fisher frequency. Let

$$T_1(t;\, n) = \left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n} \tag{2·20}$$

It can easily be shown that the rth derivative (in t) of T_1 is

$$\begin{aligned} T_1^{(r)}(t;\, n) = (-)^r \frac{(n+r-1)!}{(n-1)!\,(n-1)^r}\bigg\{&t^r - n_1\frac{r(r-1)}{2}t^{r-2} + n_2\frac{r(r-1)(r-2)(r-3)}{2\,.\,4}t^{r-4} \\ &- n_3\frac{r(r-1)(r-2)(r-3)(r-4)(r-5)}{2\,.\,4\,.\,6}t^{r-6} + \ldots\bigg\}\left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}(n+2r)}, \end{aligned} \tag{2·21}$$

with $\quad n_1 = \dfrac{n-1}{n+1}, \quad n_2 = \dfrac{(n-1)^2}{(n+1)(n+3)}, \quad n_3 = \dfrac{(n-1)^3}{(n+1)(n+3)(n+5)}, \quad$ etc.

Note that (2·21) assumes the Hermite form when $n = \infty$.

The theory will now be applied to particular examples using in all cases $n = 10$. The universes will be assumed to belong to the Karl Pearson system, so that (M. G. Kendall, 1941) the values of λ_5 and λ_6 can be derived (given λ_3 and λ_4) from the following equations:

$$\left.\begin{aligned} (1+4\eta)\,\lambda_3 + 2\xi &= 0, \\ (1+5\eta)\,\lambda_4 + 3\xi\lambda_3 + 6\eta &= 0, \\ (1+6\eta)\,\lambda_5 + 4\xi\lambda_4 + 24\eta\lambda_3 &= 0, \\ (1+7\eta)\,\lambda_6 + 5\xi\lambda_5 + 10\eta(4\lambda_4 + 3\lambda_3^2) &= 0. \end{aligned}\right\} \tag{2·22}$$

From the first two equations

$$\eta = (2\lambda_4 - 3\lambda_3^2)/(-10\lambda_4 + 12\lambda_3^2 - 12),$$

which, substituted in the first equation of (2·22), gives ξ. The values of ξ and η, substituted in the third and fourth equations, give λ_5 and λ_6. From (2·18), the L'_i being the semi-invariants

when the parent universe is normal (i.e. the values found when all the λ's are set equal to zero),

$$\left.\begin{aligned} L_1-L_1' &\simeq J_1 n^{-\frac{1}{2}}+K_1 n^{-\frac{3}{2}}, & L_4-L_4' &\simeq J_4 n^{-1}+K_4 n^{-2},\\ L_2-L_2' &\simeq J_2 n^{-1}+K_2 n^{-2}, & L_5-L_5' &\simeq \qquad\quad K_5 n^{-\frac{3}{2}},\\ L_3-L_3' &\simeq J_3 n^{-\frac{1}{2}}+K_3 n^{-\frac{3}{2}}, & L_6-L_6' &\simeq \qquad\quad K_6 n^{-2}.\end{aligned}\right\} \tag{2·23}$$

The J and K are the terms in the λ in (2·18). To n^{-2} (i.e. ignoring $n^{-\frac{5}{2}}$) the frequency generated from T of (2·19) is as follows:

$$\begin{aligned} f(t) = {} & T+n^{-\frac{1}{2}}\left\{J_1 D+\frac{J_3}{6}D^3\right\}+n^{-1}\left\{\frac{D^2}{2}(J_2+J_1^2)+\frac{D^4}{24}(J_4+4J_1J_3)+\frac{D^6}{72}J_3^2\right\}\\ & +n^{-\frac{3}{2}}\left\{K_1 D+\frac{D^3}{6}(K_3+3J_1J_2+J_1^3)+\frac{D^5}{120}(K_5+5J_1J_4+10J_2J_3+10J_1^2J_3)\right.\\ & \left.+\frac{D^7}{144}(J_3J_4+2J_1J_3^2)+\frac{J_3^3}{1296}D^9\right\}+n^{-2}\left\{\frac{D^2}{2}(K_2+2J_1K_1)+\frac{D^4}{24}(K_4+4J_1K_3\right.\\ & +4J_3K_1+3J_2^2+6J_1^2J_2+J_1^4)+\frac{D^6}{720}(K_6+6J_1K_5+20J_3K_3+15J_2J_4\\ & +60J_1J_2J_3+15J_1^2J_4+20J_1^3J_3)+\frac{D^8}{11{,}520}(16J_3K_5+10J_4^2+80J_2J_3^2\\ & \left.+80J_1J_3J_4+80J_1^2J_3^2)+\frac{D^{10}}{5184}(3J_3^2J_4+4J_1J_3^3)+\frac{J_3^4}{31{,}104}D^{12}\right\}, \end{aligned} \tag{2·24}$$

with

$$D^h=\left(-\frac{d}{dt}\right)^h T.$$

To n^{-1}, (2·24) agrees with the formula given by M. S. Bartlett (1935), in which, however, there is a small and obvious slip in a sign. The law of formation of the numerical coefficients of (2·24) is evident; for instance, the numerical coefficient of $D^8 J_2 J_3^2$ is $1/144 = 1/2!\,3!^2 2!$. The integrals $\int_t^\infty$ and $\int_{-\infty}^{-t}$ $(t>0)$ are found by reducing the exponent of D by unity, as follows:

$$\int_{-\infty}^{-t} D\,dt=-T,\quad \int_t^\infty D^{2m}\,dt=\int_{-\infty}^{-t} D^{2m}\,dt=D^{2m-1},\quad \int_t^\infty D^{2m+1}\,dt=-\int_{-\infty}^{-t} D^{2m+1}\,dt=D^{2m}. \tag{2·25}$$

In normal theory the upper and lower $2\frac{1}{2}\%$ points of t are $\pm 2{\cdot}262$ for $n=10$. Table 2 shows the 'true' probabilities, i.e. the value of

$$\int_{-\infty}^{-2\cdot 262} f(t)\,dt \tag{2·26}$$

for parent universes specified by λ_3, λ_4, using (2·24).

There are two observations to be made on the results presented in this table. The first is that, despite the considerable number of terms (shown at (2·24)) included in the probability expansion, the values found in the successive terms cannot be regarded as satisfactorily convergent for so small a sample as 10, and, of course, the convergence disimproves with increasing $\sqrt{\beta_1}$. Taken all together, however, they seem consistent and significant. The second observation is that attention was confined to the negative 'tail' of the distribution. It may be assumed that, in all cases, the distortion would be very considerably less marked if regard were had to the probability for $|t|>2{\cdot}262$. Actually for universe 3 the probability

is 0·056, not significantly different from the normal theory probability of 0·05. In justification of the attitude adopted above, the point might be put as follows:

We decide to accept the hypothesis that the universal mean is zero provided that the value of t found from the particular sample satisfies $t_0 \leqslant t \leqslant t_1$, where

$$\text{Prob}\,(t < t_0) = \text{Prob}\,(t > t_1) = 0{\cdot}025.$$

The table is designed to show that if the parent universe is markedly asymmetrical the range (t_0, t_1) may differ appreciably from $-t_0 = t_1 = 2{\cdot}262$.

Table 2. *Probabilities of t less than* $-2{\cdot}262$ *for samples of* 10 *for seven universes*

Universe	$\lambda_3 = \sqrt{\beta_1}$	$\lambda_4 = \beta_2 - 3$	Probability
Normal	0	0	0·025
2	0	1	0·024
3	1/2	0	0·041
4	$1/\sqrt{2}$	1/2	0·047
5	1	0	0·072?
6	1	1	0·086?
7	1/2	1/2	0·043

As anticipated by earlier work (*W. S. Gosset, 1908; R. C. Geary, 1936*), the table shows that the distortion is slight for symmetrical universes; even when $\lambda_4 = 1$ (and $\lambda_3 = 0$) the probability (0·024) is practically identical with the normal value. There can be little doubt that the standard table probabilities can be seriously at variance with the true probabilities when the universes from which the samples are drawn are markedly asymmetrical.

(c) *Difference of means*

R. A. Fisher's (1925) test of significance

$$t = \frac{(k_1' - k_1'')\sqrt{(n' + n'' - 2)}}{\{(n'-1)\,k_2' + (n''-1)\,k_2''\}^{\frac{1}{2}}}\sqrt{\frac{n'n''}{(n'+n'')}}, \tag{2·27}$$

for the difference of averages k_1' and k_1'' in normal theory for random samples numbering n' and n'' is, of course, a particular case of the analysis of variance considered in § (*a*) above. The second cumulants are k_2' and k_2''. It is assumed that the unknown universal means and variances are equal. Suppose now that the random samples in reality have been derived from universes in which the means are equal but the other semi-invariants λ_i' and λ_i'' are not necessarily zero for $i \geqslant 2$, or even necessarily equal. Since the universal means are assumed equal, without loss of generality we may take $\lambda_1' = \lambda_1'' = 0$. This general mathematical model seems to be the correct one; we are not trying to determine the probability of the samples being derived from the *same universe* but rather if they could conceivably have been drawn from universes with the *same arithmetic mean*, however much they may differ otherwise. The correctness or otherwise of the concept may be considered in relation to, say, the problem of deciding from two random samples which of two types of fertilizer is to be preferred from yield observations on a given crop on a given kind of land. Undoubtedly the prime problem will be that of ascertaining which is probably the better yielding (i.e. whether the arithmetic means are significantly different). Of considerably less importance is the

question of which fertilizer is the more variable; of less importance still is the question of deciding, say, whether with approximately equal yields one universe is symmetrical and the other markedly asymmetrical. The point is that the question of the equality of universal means should be considered without assuming that the other semi-invariants in the universes from which the samples have been drawn are necessarily equal. This essentially is also the viewpoint in R. A. Fisher's randomization method.

Expanding the denominator of (2·27) in terms of $(k_2'-\lambda_2')$ and $(k_2''-\lambda_2'')$ and computing therefrom the first few terms of the first four moments of t, we find the following approximations to the first four semi-invariants:

$$\left.\begin{aligned}
AL_1 &\simeq -\frac{(\lambda_3'-\lambda_3'')}{2(n'\lambda_2'+n''\lambda_2'')},\\
A^2L_2 &\simeq \left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)\left(1+\frac{2\overline{n'\lambda_2'^2+n''\lambda_2''^2}}{n'\lambda_2'+n''\lambda_2''}\right)\\
&\qquad +\frac{(n'^2-n''^2)(\lambda_4'\lambda_2''-\lambda_4''\lambda_2')}{n'n''(n'\lambda_2'+n''\lambda_2'')^2}+\frac{7(\lambda_3'-\lambda_3'')^2}{4(n'\lambda_2'+n''\lambda_2'')^2},\\
A^3L_3 &\simeq \frac{\lambda_3'}{n'^2}-\frac{\lambda_3''}{n''^2}-\frac{3(\lambda_3'-\lambda_3'')}{(n'\lambda_2'+n''\lambda_2'')}\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right),\\
A^4L_4 &\simeq \frac{6(n'\lambda_2'^2+n''\lambda_2''^2)}{(n'\lambda_2'+n''\lambda_2'')^2}\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)^2-6\left(\frac{\lambda_3'}{n'^2}-\frac{\lambda_3''}{n''^2}\right)\frac{(\lambda_3'-\lambda_3'')}{(n'\lambda_2'+n''\lambda_2'')}\\
&\quad +\frac{18(\lambda_3'-\lambda_3'')^2}{(n'\lambda_2'+n''\lambda_2'')^2}\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)+\frac{\lambda_4'}{n'^3}+\frac{\lambda_4''}{n''^3}\\
&\quad -3\left(\frac{\lambda_2'}{n'}+\frac{\lambda_2''}{n''}\right)\frac{\{\lambda_4'(n'n''\lambda_2'+\overline{2n''^2-n'^2}\lambda_2'')+\lambda_4''(n'n''\lambda_2''+\overline{2n'^2-n''^2}\lambda_2')\}}{n'n''(n'\lambda_2'+n''\lambda_2'')^2},
\end{aligned}\right\} \quad (2\cdot28)$$

with
$$A=\left\{\left(\frac{n'+n''}{n'n''}\right)\frac{(\lambda_2'\overline{n'-1}+\lambda_2''\overline{n''-1})}{(n'+n''-2)}\right\}^{\frac{1}{2}}.$$

Using formula (2·24) to the term in n^{-1} with the Gosset-Fisher function again as generating function, Table 3 shows rough approximations, for four examples, to the 'true' probability of values of $t\leqslant\tau$, where τ is the (negative) value for probability 0·025 from the normal table, and $\lambda_2'=\lambda_2''=1$. When the two samples are drawn from different universes the distortion can accordingly be considerable. The third example suggests that if the universes are the same the distortion is small, a result to be anticipated from the fact (apparent from (2·28)) that, to the approximation used, the first two semi-invariants are equal to their normal theory values; this theory confirms the experimental results of E. S. Pearson & N. K. Adyanthaya (1929).

Table 3

Example	n'	n''	λ_3'	λ_3''	λ_4'	λ_4''	Probability
1	12	4	1	−1	1	−1	0·045
2	18	6	1	−1	1	−1	0·041
3	7	4	$1/\sqrt{2}$	$1/\sqrt{2}$	1/2	1/2	0·027
4	10	6	1	0	1	0	0·036

It should be remarked that the probabilities in Table 3 (as well as in Table 2) are merely rough approximations—the samples used are far too small for the results to have any pretension to accuracy. The object has been merely to show that the actual probability *could* be considerably at variance with that shown in the standard table, for small samples.

3. Sufficient conditions for approach to normality of $a(c)$ with increasing n

The remainder of the paper deals with the field of symmetrical tests of normality, homogeneous of degree zero, represented by (3·1). It is essential to establish the conditions of approach to normality of the frequency distribution of $a(c)$ as the sample number increases.

Let
$$a(c) = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|^c \Big/ \left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{1}{2}c}, \tag{3·1}$$

where $\bar{x} = \Sigma x_i/n$ and c is non-negative. It will be shown in succession that, subject to stated conditions, with increasing n,

(i) the frequency distribution of
$$a_1(c) = \frac{1}{n}\Sigma|x_i|^c \Big/ \left\{\frac{1}{n}\Sigma x_i^2\right\}^{\frac{1}{2}c} \tag{3·2}$$
tends towards normality, and

(ii) the frequency distribution of $a_1(c)$ tends towards that of $a(c)$ and hence towards normality.

It is assumed, without loss of generality, that the universal mean of the universe from which the sample of n is drawn is zero. Denote the kth absolute moment from zero by $\mu_{|k|}$, k not being necessarily an integer. Given a positive quantity ϵ arbitrarily small, $\omega(\epsilon)$ can be found so that
$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma(|x_i|^c - \mu_{|c|})\right| < \omega\sqrt{\frac{(\mu_{|2c|}-\mu_{|c|}^2)}{n}}\right\} > 1-\epsilon, \tag{3·3}$$
$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma(x_i^2 - \mu_2)\right| < \omega\sqrt{\frac{(\mu_4-\mu_2^2)}{n}}\right\} > 1-\epsilon, \tag{3·4}$$

provided, of course, that $\mu_{|2c|}$ and μ_4 exist. As n increases ω may be envisaged as approaching the normal probability point appropriate to the probability ϵ, since, in the conditions stated, $\Sigma|x_i|^c/n$ and $\Sigma x_i^2/n$ are normally distributed in the limit. For samples which satisfy the inequality in the brackets $\{\,\}$ at (3·4) and if n is so large that
$$\omega\sqrt{\left(\frac{\mu_4}{n}\right)} < \mu_2,$$

the denominator of (3·2) can be expanded to three terms (including the remainder) by Taylor's theorem, so that $a_1(c)$ may be written
$$a_1(c) = \mu_{|c|}\mu_2^{-\frac{1}{2}c}\left\{1+\frac{1}{n}\Sigma\left(y_i - \frac{c}{2}z_i\right) - \frac{c}{2n^2}\Sigma y_i\,\Sigma z_i + \frac{c(c+2)}{8n^2}(\Sigma z_i)^2\left(1+\frac{1}{n}\Sigma y_i\right)X\right\}, \tag{3·5}$$

with
$$y_i = (|x_i|^c - \mu_{|c|})/\mu_{|c|},$$
$$z_i = (x_i^2-\mu_2)/\mu_2,$$
$$X = \mu_2^{\frac{1}{2}c+2}\left\{\mu_2 + \frac{\theta}{n}\Sigma(x_i^2-\mu_2)\right\}^{-\frac{1}{2}(c+4)} \quad (0<\theta<1).$$

With probability exceeding $(1-\epsilon)$ it is evident, from (3·4), that X is maximized by

$$\left(1-\frac{\omega}{\mu_2}\sqrt{\frac{\mu_4}{n}}\right)^{-\frac{1}{2}(c+4)}$$

It will suffice, for the present purpose, to infer that

$$|X|<\kappa,$$

where κ is a constant independent of n. We have now

$$E\frac{1}{n}\Sigma\left(y_i-\frac{c}{2}z_i\right)=0.$$

Set
$$\sigma^2 = E\frac{1}{n^2}\left\{\Sigma\left(y_i-\frac{c}{2}z_i\right)\right\}^2$$

$$=\frac{1}{n}\left\{\frac{\mu_{|2c|}}{\mu_{|c|}^2}-\frac{c\mu_{|c+2|}}{\mu_{|c|}\mu_2}+\frac{c^2}{4}\frac{\mu_4}{\mu_2^2}-\left(\frac{c}{2}-1\right)^2\right\}, \tag{3·6}$$

and
$$\frac{1}{\sigma}\left(\frac{\mu_2^{\frac{1}{2}c}a_1(c)}{\mu_{|c|}}-1\right)-\frac{1}{n\sigma}\Sigma\left(y_i-\frac{c}{2}z_i\right)=u, \tag{3·7}$$

with
$$u=-\frac{c}{2n^2\sigma}\Sigma y_i\,\Sigma z_i+\frac{c(c+2)}{8n^2\sigma}(\Sigma z_i)^2\left(1+\frac{1}{n}\Sigma y_i\right)X. \tag{3·8}$$

For samples which satisfy the inequalities in { } at (3·3) and (3·4) and hence with a probability exceeding $(1-2\epsilon)$, we have

$$|u|<\frac{c\omega^2}{2\sigma}\frac{\sqrt{(\mu_{|2c|}\mu_4)}}{n\mu_{|c|}\mu_2}+\frac{c(c+2)\kappa}{8\sigma}\frac{\omega^2\mu_4}{n\mu_2^2}\left(1+\frac{\omega}{\mu_{|c|}}\sqrt{\frac{\mu_{|2c|}}{n}}\right)<\frac{\xi}{\sqrt{n}}, \tag{3·9}$$

where ξ is independent of n. Or, briefly,

$$\text{Prob}\left\{|u|<\frac{\xi}{\sqrt{n}}\right\}>1-2\epsilon, \tag{3·10}$$

so that u tends in probability towards zero with $1/n$. Now (3·7) may be written in the form $u=Y'-Y$, where Y' and Y are the respective terms on the left side. If A be any number and F the total probability function, a well-known lemma (Fréchet, 1937, p. 164) shows that

$$|F_{Y'}(A)-F_Y(A)|\leqslant\left\{F_Y\left(A+\frac{\xi}{\sqrt{n}}\right)-F_Y\left(A-\frac{\xi}{\sqrt{n}}\right)\right\}+2\epsilon, \tag{3·11}$$

using (3·10). Hence the frequency distribution of

$$Y'=\frac{1}{\sigma}\left(\frac{\mu_2^{\frac{1}{2}c}a_1(c)}{\mu_{|c|}}-1\right) \tag{3·12}$$

tends towards that of
$$Y=\frac{1}{n\sigma}\Sigma\left(y_i-\frac{c}{2}z_i\right) \tag{3·13}$$

at every continuity point of the latter frequency, as n tends towards infinity. But Y, from (3·13), is the simple average of n random measures, and its frequency must tend towards normality provided that its standard deviation exists; from (3·6) it is evident that σ is finite provided that $\mu_{|k|}$, where k is the greater of $2c$ and 4, is finite. Here and in the remainder of this section it will be useful to remember that if $\mu_{|k|}$ exists so does $\mu_{|k'|}$ for $0\leqslant k'\leqslant k$.

To prove that the frequency distribution of $a(c)$ tends towards that of $a_1(c)$ and hence towards normality with increasing n it will be shown that $\Sigma|x_i-\bar{x}|^c/n$ tends in probability towards $\Sigma|x_i|^c/n$. Two cases will be considered separately: (1) $c\geqslant 1$, (2) $1>c\geqslant 0$.

Case (1). $c\geqslant 1$

For values of x_i for which $|x_i|\geqslant|\bar{x}|$,

$$|x_i-\bar{x}|^c-|x_i|^c=\pm c\bar{x}|x_i-\theta\bar{x}|^{c-1}\quad(0<\theta<1)$$

and when $|x_i|<|\bar{x}|$ $\qquad ||x_i-\bar{x}|^c-|x_i|^c|\leqq(2^c+1)|\bar{x}|^c.$

Hence
$$\frac{1}{n}\left|\sum_{i=1}^{n}(|x_i-\bar{x}|)^c-(|x_i|^c)\right|<|\bar{x}|\left(\frac{B}{n}\sum_{i=1}^{n}|x_i|^{c-1}+C|\bar{x}|^{c-1}\right),\tag{3·14}$$

B and C being independent of the x_i and n but depending on c. With ϵ arbitrarily small ω can be found so that

$$\left.\begin{aligned}&\text{Prob}\left\{|\bar{x}|<\omega\sqrt{\frac{\mu_2}{n}}\right\}>1-\epsilon,\\&\text{Prob}\left\{\left|\frac{1}{n}\Sigma(|x_i|^{c-1}-\mu_{|c-1|})\right|<\omega\sqrt{\frac{\mu_{|2c-2|}-\mu_{|c-1|}^2}{n}}\right\}>1-\epsilon.\end{aligned}\right\}\tag{3·15}$$

Hence, from (3·14) and (3·15), if μ_2 and $\mu_{|2c-2|}$ exist,

$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma|x_i-\bar{x}|^c-\frac{1}{n}\Sigma|x_i|^c\right|<B'\frac{\omega\mu_{|c-1|}\mu_2^{\frac{1}{2}}}{\sqrt{n}}\right\}>1-2\epsilon$$

for n sufficiently large the constant B' depending on c but not on n. Hence for $c\geqq 1$, $\Sigma|x_i-\bar{x}|^c/n$ tends in probability towards $\Sigma|x_i|^c/n$. Incidentally, this proves that $\{\Sigma(x_i-\bar{x})^2/n\}^{\frac{1}{2}c}$ tends in probability towards $\{\Sigma x_i^2/n\}^{\frac{1}{2}c}$, the latter two expressions representing respectively the denominators of $a(c)$ and $a_1(c)$.

Case (2). $1>c\geqslant 0$

Let $\bar{x}$ satisfy a probabilistic inequality identical in form with the first equation of (3·15) and let γ be any positive quantity, fixed once for all. Let n (presently to be defined further) be so large that

$$\gamma>\omega\sqrt{\frac{\mu_2}{n}}.$$

Then
$$\frac{1}{n}\sum_{i=1}^{n}(|x_i-\bar{x}|^c-|x_i|^c)=\frac{1}{n}\left(\sum_{|x_i|\geqq\gamma}{}'+\sum_{|x_i|<\gamma}{}''\right)(|x_i-\bar{x}|^c-|x_i|^c).\tag{3·16}$$

When $|x_i|\geqslant\gamma$ (i.e. in Σ'),

$$|x_i-\bar{x}|^c-|x_i|^c=\pm c\bar{x}|x_i-\theta\bar{x}|^{c-1}\quad(0<\theta<1),$$

so that
$$\text{Prob}\left\{||x_i-\bar{x}|^c-|x_i|^c|<c\omega\left(\gamma-\omega\sqrt{\frac{\mu_2}{n}}\right)^{c-1}\sqrt{\frac{\mu_2}{n}}\right\}>1-\epsilon.\tag{3·17}$$

When $|x_i|<\gamma$ (i.e. in Σ''), given η arbitrarily small and positive, n can be found so that

$$||x_i-\bar{x}|^c-|x_i|^c|<\eta,\tag{3·18}$$

when
$$|\bar{x}|<\omega\sqrt{\frac{\mu_2}{n}},$$

since $|x|^c$ $(c>0)$ is uniformly continuous in Σ''. We then have

$$\text{Prob}\{||x_i-\bar{x}|^c-|x_i|^c|<\eta\}>1-\epsilon.\tag{3·19}$$

Combining (3·17) and (3·19), it may be inferred that

$$\text{Prob}\left\{\left|\frac{1}{n}\Sigma\,|x_i-\bar{x}|^c-\frac{1}{n}\Sigma\,|x_i|^c\right|<c\omega\left(\gamma-\omega\sqrt{\frac{\mu_2}{n}}\right)^{c-1}\sqrt{\frac{\mu_2}{n}}+\eta\right\}>1-2\epsilon, \qquad (3\cdot20)$$

the first term of the upper limit in { } tending to zero as n tends towards infinity, and ϵ and η being arbitrarily small

We have accordingly shown that the numerator and denominator of $a(c)$ tends in probability towards those of $a_1(c)$. Hence $a(c)$ tends in probability towards $a_1(c)$. Hence, using the lemma cited at (3·11), the total frequency of $a(c)$ tends towards that of $a_1(c)$ which tends towards normality as n tends towards infinity. Finally:

If $c \geqslant 0$ the frequency distribution of $a(c)$, given by (3·1), *tends towards normality as n tends towards infinity provided that $\mu_{|k|}$, where k is the greater of $2c$ and 4, is finite.*

It seems likely that an analogous theorem can be proved for $0 > c > -\frac{1}{2}$; we shall not, however, be concerned in this communication with negative values of c.

4. Moments of $a(c)$ for normal samples

While it will be shown in later sections that, with indefinitely large samples, $\sqrt{b_1}$ and b_2 are the most efficient tests of asymmetry and kurtosis, respectively, it by no means follows that other tests are inefficient or that they may not be useful supplements in cases in which the prime tests are indecisive as to the probable non-normality of a given sample. It is accordingly proposed to give here close approximations to the first four moments (from the origin) of $a(c)$ (given by (3·1)) for normal random samples of n.

For normal samples (R. A. Fisher, 1929; R. C. Geary, 1933)

$$M'_k\{a(c)\} = E\{a(c)\}^k = E\left\{\frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|^c\right\}^k \Big/ E\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{1}{2}ck} \qquad (4\cdot1)$$

The exact value of the denominator is, of course, known, for

$$E\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^{\frac{1}{2}k'} = \left(\frac{n-1}{n}\right)^{\frac{1}{2}k'} Es^{k'} = \left(\frac{2}{n}\right)^{\frac{1}{2}k'}\left(\frac{n+k'-3}{2}\right)! \Big/ \left(\frac{n-3}{2}\right)!, \qquad (4\cdot2)$$

since, as usual, $(n-1)s^2 = \Sigma(x_i-\bar{x})^2$. It will be useful to expand $\log_e Es^{k'}$ with $k' = ck$ using Stirling's formula in (4·2):

$$\begin{aligned}\log_e Es^{k'} &= \frac{k'}{2}\log\frac{2}{n-1}+\log\left(\frac{n+k'-3}{2}\right)!-\log\left(\frac{n-3}{2}\right)!\\ &= \frac{(k'^2-2k')}{4(n-1)}-\frac{k'(k'-1)(k'-2)}{12(n-1)^2}+\frac{k'^2(k'-2)^2}{24(n-1)^3}-\frac{k'(k'-1)(k'-2)(3k'^2-6k'-4)}{120(n-1)^4}\\ &\quad+\frac{k'^2(k'-2)^2(k'^2-2k'-2)}{60(n-1)^5}-\frac{k'(k'-1)(k'-2)(3k'^4-12k'^3+24k'+16)}{252(n-1)^6}\\ &\quad+\frac{k'^2(k'-2)^2(3k'^4-12k'^3-4k'^2+32k'+32)}{336(n-1)^7},\end{aligned} \qquad (4\cdot3)$$

which checks for $k'=1$ to $(n-1)^{-7}$ with Geary (1935, p. 354). Take

$$v(c) = \frac{1}{n}\sum_{i=1}^{n}|z_i|^c, \qquad (4\cdot4)$$

with

$$z_i = x_i - \bar{x}.$$

The moments of $v(c)$ will be found exactly as in the case of $c = 1$ (Geary, 1936) from the single or joint normal frequency distributions of $(z_1, z_2, \ldots)$. We find

$$M_1'\{v(c)\} = \frac{1}{\sqrt{\pi}}\left(\frac{2\,\overline{n-1}}{n}\right)^{\frac{1}{2}c}\left(\frac{c-1}{2}\right)!,$$

$$M_2'\{v(c)\} = \frac{1}{\sqrt{\pi}}\frac{(2\,\overline{n-1})^c}{n^{c+1}}\left(\frac{2c-1}{2}\right)! + \frac{2^c}{\pi}n^{-\frac{1}{2}}(n-1)^{-c}(n-2)^{\frac{1}{2}(2c+1)}\left[\left(\frac{c-1}{2}\right)!\right]^2 \tag{4·5}$$

$$\times\left\{1+\frac{1}{2!}\left(\frac{c+1}{2}\right)^2\left(\frac{2}{n-1}\right)^2+\frac{1}{4!}\left(\frac{c+1}{2}\right)^2\left(\frac{c+3}{2}\right)^2\left(\frac{2}{n-1}\right)^4+\frac{1}{6!}\left(\frac{c+1}{2}\right)^2\left(\frac{c+3}{2}\right)^2\left(\frac{c+5}{2}\right)^2\left(\frac{2}{n-1}\right)^6+\ldots\right\}. \tag{4·6}$$

For the third moment we write

$$M_3'\{v(c)\} = E\{v(c)\}^3 = \frac{n}{n^3}E\,|\,z_1\,|^{3c} + \frac{3n(n-1)}{n^3}E\,|\,z_1\,|^{2c}\,|\,z_2\,|^c + \frac{n(n-1)(n-2)}{n^3}E\,|\,z_1\,|^c\,|\,z_2\,|^c\,|\,z_3\,|^c$$

$$= A_1 + A_2 + A_3, \tag{4·7}$$

denoting the three terms on the right by A_1, A_2, A_3 respectively. Then

$$A_1 = \frac{1}{\sqrt{\pi}}\left(\frac{3c-1}{2}\right)!(2\,\overline{n-1})^{\frac{3}{2}c}\,n^{-\frac{1}{2}(4+3c)},$$

$$A_2 = \frac{3\,.\,2^{\frac{3}{2}c}}{\pi}\left(\frac{2c-1}{2}\right)!\left(\frac{c-1}{2}\right)!(n-2)^{\frac{1}{2}(3c+1)}(n-1)^{-\frac{3}{2}c}n^{-\frac{3}{2}}$$

$$\times\left\{1+\frac{(2c+1)(c+1)}{2!(n-1)^2}+\frac{(2c+3)(2c+1)(c+3)(c+1)}{4!(n-1)^4}+\ldots\right\},$$

$$A_3 = \left(\frac{2^c}{\pi}\right)^{\frac{3}{2}}(n-3)^{\frac{1}{2}(3c+2)}(n-2)^{-\frac{1}{2}(3c+1)}(n-1)\,n^{-\frac{3}{2}}\left[\left(\frac{c-1}{2}\right)!\right]^3\left\{1+\frac{3}{2}\frac{(c+1)^2}{(n-2)^2}\right.$$

$$-\frac{(c+1)^3}{(n-2)^3}+\frac{(c+1)^2(c+3)(7c+9)}{8(n-2)^4}-\frac{(c+3)^2(c+1)^3}{2(n-2)^5}$$

$$\left.+\frac{(c+3)^2(c+1)^2(61c^2+310c+265)}{240(n-2)^6}+\ldots\right\}.$$

Similarly, for the fourth moment,

$$M_4'\{v(c)\} = E\{v(c)\}^4 = \frac{n}{n^4}E\,|\,z_1\,|^{4c}+\frac{4n(n-1)}{n^4}E\,|\,z_1\,|^{3c}\,|\,z_2\,|^c$$

$$+\frac{3n(n-1)}{n^4}E\,|\,z_1\,|^{2c}\,|\,z_2\,|^{2c}+\frac{6n(n-1)(n-2)}{n^4}E\,|\,z_1\,|^{2c}\,|\,z_2\,|^c\,|\,z_3\,|^c$$

$$+\frac{n(n-1)(n-2)(n-3)}{n^4}E\,|\,z_1\,|^c\,|\,z_2\,|^c\,|\,z_3\,|^c\,|\,z_4\,|^c$$

$$= C_1+C_2+C_3+C_4+C_5 \tag{4·8}$$

with

$$C_1 = \frac{2^{2c}}{\sqrt{\pi}}(n-1)^{2c}n^{-2c-3}\left(\frac{4c-1}{2}\right)!,$$

$$C_2 = \frac{2^{2(c+1)}}{\pi}(n-2)^{\frac{1}{2}(4c+1)}(n-1)^{-2c}n^{-\frac{3}{2}}\left(\frac{3c-1}{2}\right)!\left(\frac{c-1}{2}\right)!$$

$$\times\left\{1+\frac{(3c+1)(c+1)}{2!(n-1)^2}+\frac{(3c+3)(3c+1)(c+3)(c+1)}{4!(n-1)^4}+\ldots\right\},$$

$$C_3 = \frac{3\,.\,2^{2c}}{\pi}(n-2)^{\frac{1}{2}(4c+1)}(n-1)^{-2c}n^{-\frac{1}{2}}\left[\left(\frac{2c-1}{2}\right)!\right]^2\left\{1+\frac{(2c+1)^2}{2!(n-1)^2}+\frac{(2c+3)^2(2c+1)^2}{4!(n-1)^4}+\dots\right\},$$

$$C_4 = \frac{3\,.\,2^{2c+1}}{\pi^{\frac{3}{2}}}(n-3)^{2c+1}(n-2)^{-\frac{1}{2}(4c+1)}(n-1)\,n^{-\frac{1}{2}}\left(\frac{2c-1}{2}\right)!\left[\left(\frac{c-1}{2}\right)!\right]^2$$

$$\times\left\{1+\frac{(c+1)(5c+3)}{2(n-2)^2}-\frac{(c+1)^2(2c+1)}{(n-2)^3}+\frac{(c+1)(57c^3+227c^2+255c+81)}{24(n-2)^4}\right.$$

$$\left.-\frac{(2c+1)(c+1)^2(c+3)(5c+9)}{6(n-2)^5}+\dots\right\},$$

$$C_5 = \frac{2^{2c}}{\pi^2}(n-4)^{\frac{1}{2}(4c+5)}(n-3)^{-2c-1}(n-2)(n-1)\,n^{-\frac{1}{2}}\left[\left(\frac{c-1}{2}\right)!\right]^4$$

$$+\left\{1+\frac{3(c+1)^2}{(n-3)^2}-\frac{4(c+1)^3}{(n-3)^3}+\frac{(c+1)^2(7c^2+21c+15)}{(n-3)^4}-\frac{4(c+3)(c+1)^3(2c+3)}{(n-3)^5}\right.$$

$$\left.+\frac{(c+3)(c+1)^2(122c^3+671c^2+1070c+525)}{15(n-3)^6}-\dots\right\}.$$

Formulae (4·5), (4·6), (4·7) and (4·8) were checked from the corresponding formulae for $c = 1$ given in the author's 1936 paper.

From the following section it will be apparent that for indefinitely large samples the most sensitive test of kurtosis of the field $a(c)$ is found for $c = 4$. At the same time it is shown that there is really not much difference in efficiency for values of c in the range $5 \geqslant c > 2$; moreover, the results in § 6 (in which the efficiency of the tests for $c = 4$ and $c = 1$ are compared from the power function viewpoint) suggest that, for samples of moderate size, the superiority, if any at all, of a test using $a(4) = b_2$ over other tests in the series may be even less marked. The disadvantage of $a(4)$ is that its frequency is not known for samples of all sizes; and if we could estimate, with any degree of confidence, the probability points of $a(c)$ for any value or values of $c > 2$ for medium-size samples we might, for practical purposes, dispense with $a(4)$ altogether, since, while we now know one way of solving the problem of determining the exact, or almost exact, frequency distribution of $a(4)$, it must be admitted that the method is extremely tedious. (From the theoretical point of view, however, the $a(4)$ problem must be solved since it remains a challenge to the mathematical skill of statisticians!) It will accordingly be of interest to study the order of magnitude of the semi-invariants of $a(c)$ for c near 2.

Consider the case, for example, of $c = 2{\cdot}4$, not by any means, it is important to observe, the lowest value which would be used for tabulating. In Table 4 the first three moments are given for $n = 25$. The L's represent, of course, the semi-invariants. The values of the functions for $a_1(c)$ (given by (3·2)) for $n = 24$ (i.e. the appropriate number of degrees of freedom for comparison with $a(c)$) are also given. These show that the moments of $a_1(c)$ are very close to those of $a(c)$, which suggests that, when n is not less than, say, 20, the values of B_1, B_2 and corresponding functions of higher orders, if required, for $a_1(c)$ could be used for the determination of the probability points of $a(c)$. This is important from the computational point of view because the algebraic expressions *for the normal moments of* $a_1(c)$ are exceedingly simple whereas it must be conceded that (4·8) offers a grim prospect for the computer; furthermore, the principal term C_5 is rather slowly convergent unless $n > 50$ or so,

whereas *exact* values for all values of n can readily be found for the moments of $a_1(c)$ for normal samples.

Table 4. *Normal moments, etc., of $a(c)$ and $a_1(c)$ for $c = 2\cdot4$*

	$a(2\cdot4)$	$a_1(2\cdot4)$
n	25	24
$M_1' = L_1$	1·166252	1·1662524891
M_2'	1·362004	1·362091186
M_3'	1·592841	1·593151615
$M_2 = L_2$	0·001860	0·001946318
$M_3 = L_3$	0·000063	0·000069583
$\sqrt{B_1} = L_3/L_2^{\frac{3}{2}}$	0·80	0·8104

As with (4·1) for $a(c)$, the moments (from the origin) of any order of $a_1(c)$ is the quotient of the moments of the same order for numerator and denominator, assuming that the universal mean is zero and the variance unity. Since the different members x_i of the sample are independent—the difficulty with $a(c)$ is that the $(x_i - \bar{x})$ are *not* independent—for the moments of the numerator of (3·2) we require only

$$E\,|\,x\,|^{k'} = \frac{2}{\sqrt{(2\pi)}}\int_0^\infty dx\, x^{k'}\, e^{-\frac{1}{2}x^2} = \left(\frac{k'-1}{2}\right)!\,\frac{2^{\frac{1}{2}k'}}{\sqrt{\pi}}, \tag{4·9}$$

and for the denominator

$$Es^{k'} = E\left(\frac{1}{n}\sum_i x_i^2\right)^{\frac{1}{2}k} = \left(\frac{2}{n}\right)^{\frac{1}{2}k'}\left(\frac{n+k'-2}{2}\right)!\Big/\left(\frac{n-2}{2}\right)!. \tag{4·10}$$

The case of $c = 4$ is particularly simple. The first four semi-invariants are as follows:

$$\left.\begin{aligned}
L_1 &= M_1' = \frac{3n}{(n+2)},\\
L_2 &= M_2 = \frac{24n^2(n-1)}{(n+2)^2(n+4)(n+6)},\\
L_3 &= M_3 = \frac{1728(n-1)(n-2)n^3}{(n+2)^3(n+4)(n+6)(n+8)(n+10)},\\
L_4 &= \frac{10{,}368n^4(n-1)(30n^4+168n^3-608n^2-2672n+3712)}{(n+2)^4(n+4)^2(n+6)^2(n+8)(n+10)(n+12)(n+14)}.
\end{aligned}\right\} \tag{4·11}$$

Moments, etc., for $a_1(c)$ for normal samples of 24 and 50 are contrasted for $c = 2\cdot4$ and $c = 4$ in Table 5. The contrast between the values of $\sqrt{B_1}$ and $(B_2 - 3)$ respectively for $a_1(2\cdot4)$ and $a_1(4)$ is striking in the extreme. Even for $n = 24$ $\sqrt{B_1}\,[a_1(2\cdot4)]$ and $B_2\,[a_1(2\cdot4)]$ are approaching the values at which a Gram-Charlier approximation to the frequency distribution may be reasonably convergent. Furthermore, the decline in the values of the B's from $n = 24$ to $n = 50$ is marked for $a_1(2\cdot4)$, while the decline in the $B[a_1(4)]$ is very slow.

It is accordingly suggested that a table of probability points (perhaps 0·001, 0·01, 0·025, 0·05 and 0·10) of $a(c)$, for c equal to, say, 2·2, be prepared for $n \geqslant 25$ on the assumption that Gram-Charlier applies throughout. For this purpose the values of the mean and variance for n at intervals of, say, 10 should be computed from formulae (4·5) and (4·6); the B_1 and $(B_2 - 3)$ should, however, be computed as for $a_1(c)$. For lower sample sizes it might be well

to use terms to order n^{-2} which would render necessary the use of the fifth and sixth semi-invariants of $a_1(c)$. The formulae given by E. A. Cornish & R. A. Fisher (1937) (assuming Gram-Charlier) could be used to find the probability points. On account of the minuteness of the variance L_2 for c near 2 it will be necessary to work to many places of decimals—at least 10. As stated at the outset, the test of kurtosis $a(2{\cdot}2)$ will be only slightly less efficient than $a(4)$ and it may be slightly more efficient than $a(1)$, the probability points of which are known approximately for samples of all sizes. In any case the $a(2{\cdot}2)$ table would be a useful adjunct to that of $a(1)$.

Table 5. *Normal moments, etc., of $a_1(c)$ for $c = 2{\cdot}4$ and $c = 4$*

	$n=24$		$n=50$	
	$c=2{\cdot}4$	$c=4$	$c=2{\cdot}4$	$c=4$
$M_1'=L_1$	1·1662524891	2·769231	1·1721603127	2·884615
$M_2=L_2$	0·001946318	0·559932	0·001058462	0·359550
$M_3=L_3$	0·000069583	0·752488	0·000022251	0·343337
$L_4=M_4-3L_2^2$	0·000004921	1·955999	0·000000919	0·711375
$\sqrt{B_1}=L_3/L_2^{\frac{3}{2}}$	0·8104	1·7960	0·6462	1·5925
$B_2-3=L_4/L_2^2$	1·30	6·24	0·82	5·50

In an earlier paper (1935) the writer suggested that the correlation between b_2 and $a(1)$ for normal samples gave some indication of the relative efficiency of these two tests of normality. In this order of ideas it seems desirable to compute the approximate value of the correlation coefficient between $a(c)$ and $a(c')$, where c and c' are any two positive constants. In the first instance the universe from which the sample of n was drawn was not necessarily normal. Since in the present application we will be concerned only with large samples we assume the universal mean known (and accordingly it may be taken as zero, i.e. $\lambda_1 = 0$), so that, instead of $a(c)$ we use, in reality, $a_1(c)$ given by (3·2). In the remainder of this section we write a for $a_1(c)$ and a' for $a_1(c')$:

$$a = \left(\frac{1}{n}\Sigma\,|\,x_i\,|^c\right)\Big/\left(\frac{1}{n}\Sigma x_i^2\right)^{\frac{1}{2}c}, \tag{4·12}$$

$$a' = \left(\frac{1}{n}\Sigma\,|\,x_i\,|^{c'}\right)\Big/\left(\frac{1}{n}\Sigma x_i^2\right)^{\frac{1}{2}c'}. \tag{4·13}$$

Set
$$\left.\begin{aligned} y_i &= (|\,x_i\,|^c-\mu_{|c|})/\mu_{|c|},\\ y_i' &= (|\,x_i\,|^{c'}-\mu_{|c'|})/\mu_{|c'|},\\ z_i &= (x_i^2-\mu_2)/\mu_2,\\ \alpha &= \mu_{|c|}/\mu_2^{\frac{1}{2}c}, \quad \alpha' = \mu_{|c'|}/\mu_2^{\frac{1}{2}c'},\\ C &= \frac{c+c'}{2}, \quad C_k = \frac{C(C+1)(C+2)\dots(C+k-1)}{k!}. \end{aligned}\right\} \tag{4·14}$$

Then
$$\frac{aa'}{\alpha\alpha'} = \left(1+\frac{1}{n}\Sigma y_i\right)\left(1+\frac{1}{n}\Sigma y_i'\right)\left(1+\frac{1}{n}\Sigma z_i\right)^{-\frac{1}{2}(c+c')} \tag{4·15}$$

The mean value of $aa'/\alpha\alpha'$ was found approximately (i.e. to terms in n^{-3}) by formally expanding the last factor in (4·15), multiplying by the first two factors, and setting down the mean value term by term, so that

$$
\begin{aligned}
M'_{cc'}/\alpha\alpha' = Eaa'/\alpha\alpha' \simeq \Big\{ & 1+\frac{1}{n^2}C_2 nEz^2 - \frac{1}{n^3}C_3 nEz^3 \\
& + \frac{1}{n^4}C_4\left(nEz^4+\frac{6n\overline{n-1}}{2}E^2z^2\right) - \frac{1}{n^5}C_5(10n\overline{n-1}\,Ez^3Ez^2) \\
& + \frac{1}{n^6}C_6\,90\,\frac{n\overline{n-1}\,\overline{n-2}}{6}E^3z^2\Big\} + \Big\{-\frac{C_1}{n^2}(nEyz+nEy'z)+\frac{C_2}{n^3}(nEyz^2+nEy'z^2) \\
& - \frac{C_3}{n^4}[n(Eyz^3+Ey'z^3)+3n\overline{n-1}\,Ez^2(Eyz+Ey'z)] \\
& + \frac{C_4}{n^5}[4n\overline{n-1}\,Ez^3(Eyz+Ey'z)+6n\overline{n-1}\,Ez^2(Eyz^2+Ey'z^2)] \\
& - \frac{30C_5}{n^6}\frac{n\overline{n-1}\,\overline{n-2}}{2}E^2z^2(Eyz+Ey'z)\Big\} + \Big\{\frac{1}{n^2}nEyy' - \frac{C_1}{n^3}nEyy'z \\
& + \frac{C_2}{n^4}[nEyy'z^2+2n\overline{n-1}\,Eyz\,Ey'z+n\overline{n-1}\,Eyy'\,Ez^2] \\
& - \frac{C_3}{n^5}[n\overline{n-1}\,Eyy'\,Ez^3+3n\overline{n-1}\,(Eyz^2Ey'z+Ey'z^2Eyz+Eyy'zEz^2)] \\
& + \frac{C_4}{n^6}\left[\frac{6n\overline{n-1}\,\overline{n-2}}{2}Eyy'E^2z^2+12n\overline{n-1}\,\overline{n-2}\,EyzEy'zEz^2\right]\Big\}.
\end{aligned} \qquad (4\cdot16)
$$

The E's in (4·16) are readily calculable from (4·14), e.g.

$$Eyy' = Ey_iy'_i = E(|x_i|^c-\mu_{|c|})(|x_i|^{c'}-\mu_{|c'|})/\mu_{|c|}\mu_{|c'|} = (\mu_{|c+c'|}/\mu_{|c|}\mu_{|c'|})-1.$$

It has been verified that when c is substituted for c' in (4·13) the formula agrees with that for the second moment of $a_1(c)$ given in § 6.

The coefficient of correlation is, of course,

$$R_{cc} = M_{cc'}/\sqrt{(M_{cc}M_{c'c'})}, \qquad (4\cdot17)$$

with

$$M_{cc'} = M'_{cc'} - M'_c M'_{c'}.$$

Formulae for the first and second moments, to the approximation required, for the computation of (4·17) are given in § 6.

As an application, the following are the values of the variances and the covariance for the test of normality $a(1)$ and (b_2), i.e. in which c and c' have respectively the values 1 and 4, and where the universe belongs to the Pearson system with $\lambda_2 = 1$, $\lambda_3 = 0$ and $\lambda_4 = \frac{1}{2}$:

$$
\left.
\begin{aligned}
\frac{M_{cc}}{\mu^2_{|c|}} &\simeq \frac{0\cdot09313705}{n} - \frac{0\cdot262961}{n^2} - \frac{0\cdot196477}{n^3}, \\
\frac{M_{c'c'}}{\mu^2_{|c'|}} &\simeq \frac{4\cdot4286}{n} - \frac{92\cdot25}{n^2} + \frac{831\cdot2}{n^3}, \\
\frac{M_{cc'}}{\mu_{|c|}\mu_{|c'|}} &\simeq -\frac{0\cdot491}{n} + \frac{4\cdot87}{n^2} - \frac{281\cdot5}{n^3}.
\end{aligned}
\right\} \qquad (4\cdot18)
$$

From (4·17) and (4·18), $R_{cc'}(n=100) \simeq -0·826$ and $R_{cc'}(n=\infty) = -0·764$. It is of great interest to find that, though the universe is markedly non-normal the correlation for indefinitely large samples is practically identical with the normal theory value of $-0·767$ (Geary, 1935), another indication, no doubt, that normal theory inferences can usually be applied with confidence when the parent universe is not markedly unsymmetrical.

When samples are indefinitely large we find, from (4·16) and (4·17),

$$R_{cc'} = \frac{4\mu_{|c+c'|} - 2(c\mu_{|c|}\mu_{|c'+2|} + c'\mu_{|c'|}\mu_{|c+2|}) + (cc'\mu_4 - \overline{c-2}\,.\,\overline{c'-2})\,\mu_{|c|}\mu_{|c'|}}{\sqrt{(M_{cc}M_{c'c'})}}, \tag{4·19}$$

where, of course, the values to be taken here for M_{cc} and $M_{c'c'}$ are found by substituting respectively c' for c and c for c' in the numerator. When, in addition, the parent universe is normal, we find

$$R^0_{cc'} = \frac{\left(\frac{c+c'-1}{2}\right)!\sqrt{\pi} - \left(\frac{c-1}{2}\right)!\left(\frac{c'-1}{2}\right)!\left(\frac{cc'+2}{2}\right)}{\sqrt{\left[\left\{\left(\frac{2c-1}{2}\right)!\sqrt{\pi} - \left(\frac{c-1}{2}!\right)^2\left(\frac{c^2+2}{2}\right)\right\}\left\{\left(\frac{2c'-1}{2}\right)!\sqrt{\pi} - \left(\frac{c'-1}{2}!\right)^2\left(\frac{c'^2+2}{2}\right)\right\}\right]}} \tag{4·20}$$

which reduces to $-1/\sqrt{\{12(\pi-3)\}}$ for $c = 1$, $c' \doteq 4$, as it should (Geary, 1935). The following section will accord b_2 (i.e. $a(4)$) a decided primacy amongst tests of normality when the samples are indefinitely large. It may, therefore, be of interest to give the values of the correlation coefficients (for indefinitely large normal samples) between b_2 and $a(c)$ for selected values of c (Table 6). The table suggests, in the high coefficients of correlation, except for c very near 0 or 2, that all the $a(c)$ should be reliable tests of kurtosis, with no great difference between their efficiencies. The efficiency of any two tests would be identical, in the conditions stated, if the coefficient of correlation between them was ± 1 because then, of course, they would be functionally, and not stochastically, related.

Table 6. *Correlation between b_2 and $a(c)$ for indefinitely large normal samples*

Value of c	Value of R^0_{c4}	Value of c	Value of R^0_{c4}
0	0	3	0·980
1	$-0·769$	4	1
2	0	5	0·983
2·2	0·887	6	0·939
2·5	0·952	∞	0

5. The most efficient tests for indefinitely large samples

In this section we consider the efficiency of tests of kurtosis and asymmetry from the viewpoint of indefinitely large samples.

By definition a test will be regarded as *valid*, in relation to a field of continuous alternative universes including the normal, if its value for infinite samples drawn at random from the normal universe is different from its value for infinite samples from other universes of the field. As the sample number increases the test will become increasingly discriminatory of the normal as distinct from other universes of the field. This increased sensitivity might be given mathematical expression in some such terms as the following: given a probability α (say 0·01), the normal universe W_0 of the field and any other distribution W_1 of the field,

a number n_1 can be found so that for $n \geqslant n_1$ the mean value of the test function for samples of n from W_1 will lie at or beyond the α probability point of the test function for samples of n from W_0: the smaller n_1 the more sensitive the test.

We consider, then, the infinite field of alternative tests of kurtosis represented by (3·1) when c assumes all positive values, and the infinite field of alternative universes represented by the Gram-Charlier frequency

$$\frac{1}{\sqrt{(2\pi)}}\exp\left\{\sum_{i=3}^{\infty}\frac{\lambda_i}{i!}\left(-\frac{d}{dx}\right)^i\right\}e^{-\frac{1}{2}x^2}. \tag{5·1}$$

The universal variance is assumed to be unity, without loss of generality. The normal universe is a member of the field: it is found when all the λ_i $(i > 2)$ are zero. We assume that the conditions of § 2 are satisfied so that for indefinitely large samples the frequency distribution of $a(c)$ for all parent universes is normal. Obviously the efficiency of any particular test (i.e. $a(c)$ for a particular value of c) in regard to the normal and a particular non-normal alternative (i.e. a Gram-Charlier frequency with particular values of the λ_i) will be adjudged by considering the ratio of

(i) the difference between the universal mean values of $a(c)$ for the normal and the particular non-normal parent universes; to

(ii) the standard deviation of $a(c)$ for indefinitely large normal samples.

The most efficient test will be $a(c)$ for c a theoretically ascertainable function of the given λ_i which makes the ratio a maximum.

For indefinitely large samples the mean value ϕ of $a(c)$ when the parent universe is given by (5·1) is

$$\phi = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} dx\,|\,x\,|^c \exp\left\{\sum_i \frac{\lambda_i}{i!}\left(-\frac{d}{dx}\right)^i\right\}e^{-\frac{1}{2}x^2}. \tag{5·2}$$

Obviously
$$\int_{-\infty}^{\infty} dx\,|\,x\,|^c\left(-\frac{d}{dx}\right)^{2m+1} e^{-\frac{1}{2}x^2} = 0.$$

Also, when $m \geqslant 1$,

$$\int_{-\infty}^{\infty} dx\,|\,x\,|^c\left(\frac{d}{dx}\right)^{2m} e^{-\frac{1}{2}x^2} = \left(\frac{c-1}{2}\right)!\,2^{\frac{1}{2}(c+1)}c(c-2)\,(c-4)\ldots(c-2m+2), \tag{5·3}$$

a result readily inferable from the obvious fact that the left side vanishes for $c = 0, 2, \ldots, 2m-2$. Accordingly

$$\phi = \left(\frac{c-1}{2}\right)!\,\frac{2^{\frac{1}{2}(c+1)}}{\sqrt{(2\pi)}}\left\{1+\frac{\lambda_4}{24}c(c-2)+\left(\frac{\lambda_3^2}{72}+\frac{\lambda_6}{720}\right)c(c-2)\,(c-4)+\ldots\right\}. \tag{5·4}$$

The normal value is given by the first term.

From (4·3), (4·5) and (4·6) it is evident that the value of the standard deviation, for larger normal samples (retaining only $n^{-\frac{1}{2}}$) is

$$\sigma = \frac{2^{\frac{1}{2}c}}{\sqrt{(\pi n)}}\left(\left(\frac{2c-1}{2}\right)!\sqrt{\pi}-\left(\frac{c-1}{2}!\right)^2\frac{c^2+2}{2}\right)^{\frac{1}{2}}. \tag{5·5}$$

The principal term in the deviation $\phi-\phi^0$ (where ϕ^0 is the normal value), from (5·4), is

$$\delta = \frac{\frac{1}{2}(c-1)!\,2^{\frac{1}{2}c}}{\sqrt{\pi}}\cdot\frac{\lambda_4 c(c-2)}{24}. \tag{5·6}$$

To a constant factor, the ratio δ/σ is given by the *first discriminant*

$$\rho(c) = c(c-2)\left\{\frac{\left(\frac{2c-1}{2}\right)!\sqrt{\pi}}{\left(\frac{c-1}{2}!\right)^2} - \frac{c^2+2}{2}\right\}^{-\frac{1}{2}} \tag{5·7}$$

It will now be shown that $\dfrac{d\rho(c)}{dc} = 0$ for $c = 4$.

The discriminant may be written in the form

$$\rho(c) = c(c-2)\left(\frac{2^c I_{2c}}{I_c} - \frac{c^2+2}{2}\right)^{-\frac{1}{2}}, \tag{5·8}$$

where

$$I_c = \int_0^{\frac{1}{2}\pi} \cos^c\theta\, d\theta, \tag{5·9}$$

and

$$\frac{\rho'(c)}{\rho(c)} = \frac{1}{c} + \frac{1}{c-2} - \frac{1}{2}\left\{2^c\left(\frac{I_{2c}\log 2}{I_c} + \frac{I'_{2c}}{I_c} - \frac{I_{2c}I'_c}{I_c^2}\right) - c\right\}\Big/\left\{\frac{2^c I_{2c}}{I_c} - \frac{c^2+2}{2}\right\}. \tag{5·10}$$

From (5·9)

$$J_c = I'_c = \int_0^{\frac{1}{2}\pi} d\theta \log^c\theta \log\cos\theta. \tag{5·11}$$

From a fairly well-known property

$$J_0 = \int_0^{\frac{1}{2}\pi} d\theta \log\cos\theta = -\tfrac{1}{2}\pi\log 2. \tag{5·12}$$

In (5·10) we shall be concerned only with even positive integer values of c. We have at once

$$I_0 = \frac{\pi}{2},\quad I_2 = \frac{\pi}{4},\quad I_4 = \frac{3\pi}{16},\quad I_6 = \frac{5\pi}{32},\quad I_8 = \frac{35\pi}{256}. \tag{5·13}$$

From (5·11) $\quad J_{2c} = \displaystyle\int_0^{\frac{1}{2}\pi} d\theta \cos^{2c}\theta \log\cos\theta = \int_0^{\frac{1}{2}\pi} d(\sin\theta)\cos^{2c-1}\theta\log\cos\theta,$

which, by partial integration,

$$= \int_0^{\frac{1}{2}\pi} d\theta \sin\theta\left(\overline{2c-1}\sin\theta\cos^{2c-2}\theta\log\cos\theta + \frac{\cos^{2c-1}\theta\sin\theta}{\cos\theta}\right)$$

$$= (2c-1)(J_{2c-2} - J_{2c}) + I_{2c-2} - I_{2c}.$$

Hence

$$2cJ_{2c} = (2c-1)J_{2c-2} - I_{2c} + I_{2c-2}. \tag{5·14}$$

From (5·12), (5·13) and (5·14),

$$\left.\begin{aligned} J_0 &= -\tfrac{1}{2}\pi\log 2, & J_6 &= (-60\pi\log 2 + 37\pi)/384,\\ J_2 &= (-2\pi\log 2 + \pi)/8, & J_8 &= (-840\pi\log 2 + 533\pi)/6144.\\ J_4 &= (-12\pi\log 2 + 7\pi)/64, & & \end{aligned}\right\} \tag{5·15}$$

Noting that $I'_{2c} = 2J_{2c}$ and substituting in the right side of (5·10) the values of I and J given by (5·13) and (5·15), we find $\rho'(4) = 0$. Table 7 gives the values of the discriminant for certain values of c.

The discriminant accordingly assumes a maximum value for $c = 4$, a result so remarkable that one might be inclined to suspect that it is a consequence of the form which was assumed for the alternative to the normal curve, a form which, in placing such emphasis on λ_4,

high-lights, so to speak, b_2 ($= \lambda_4 + 3$ when $\lambda_2 = 1$ for indefinitely large samples) as a test of normality. From the algebraic point of view this is anything but obvious: the property emerges from quite a complicated piece of algebra. It may also be emphasized that the field of alternatives (5·1) is not arbitrary; it is a general form of frequency distribution when all the λ_i are finite. Admittedly the discriminant takes account only of the term in λ_4 in the expansion; but this is certainly the most significant term for a wide class of frequency distributions, namely, those of homogeneous symmetrical functions of samples of n as n tends towards infinity under very general conditions for the parent universe, provided that the resulting frequency distribution can be assumed to have its third moment zero; for then the only term in n^{-1} in the frequency distribution of the function will be the term in λ_4. The significance of the property demonstrated must not be overstressed since it is subject to many qualifications, but it gives strong grounds for holding that, for very large samples, b_2 is the most efficient test of normality of tests of type $a(c)$ in relation to a very extended class of alternative universes. At the same time Table 7 shows that there can be little difference in efficiency in the field $a(c)$ for c ranging from close to 2 to about 5. There is but little doubt, on this showing, that b_2 is more sensitive than $a(1)$, a conclusion suggested on the basis of certain experimental results by E. S. Pearson (1935) and examined from the viewpoint of power function theory in § 6.

Table 7

$0<c<2$	Discriminant $\rho(c)$	$2<c<\infty$	Discriminant $\rho(c)$
+0	−2·334	2+0	4·460
0·1	−2·541	2·1	4·508
0·2	−2·725	2·5	4·666
0·5	−3·188	3·0	4·801
0·7	−3·441	3·9	4·898
1·0	−3·758	**4·0**	**4·900**
1·1	−3·851	4·1	4·898
1·5	−4·166	5·0	4·818
1·9	−4·405	6·0	4·602
2−0	−4·460	7·0	4·288
		8·0	3·906

Adverting to (5·4) in conjunction with (5·5), it might be asked if, on the analogy of the maximal property just demonstrated for the first discriminant, the function

$$\rho_2(c) = c(c-2)(c-4)\left\{\frac{\left(\frac{2c-1}{2}\right)!\sqrt{\pi}}{\left(\frac{c-1}{2}!\right)^2} - \frac{c^2+2}{2}\right\}^{-\frac{1}{4}}$$

has a turning point at $c = 6$. The answer is in the negative. The value of $\rho_2'(6)/\rho_2(6)$ is, in fact, 15/34. At the same time there must be a zero of $\rho_2'(c)$ very near $c = 6$ since

$$\rho_2(5{\cdot}9) = 8{\cdot}79, \quad \rho_2(6) = 9{\cdot}20, \quad \rho_2(6{\cdot}1) = 8{\cdot}56.$$

Analogous to the field on tests of kurtosis represented by (3·1) we may consider as a field of tests of asymmetry:

$$g(c) = \frac{1}{n}\{-\Sigma' \mid x_i - \bar{x} \mid^c + \Sigma''(x_i - \bar{x})^c\} \Big/ \left\{\frac{1}{n}\Sigma(x_i - \bar{x})^2\right\}^{\frac{1}{2}c}, \qquad (5{\cdot}16)$$

where Σ' extends to the observations x_i less than the mean $\bar{x}$ and Σ'' to the rest of the sample. For $c = 3$ the test is, of course, $\sqrt{b_1}$. For normal samples

$$E\{g(c)\}^k = E\left\{-\frac{1}{n}\Sigma'\,|\,x_i-\bar{x}\,|^c+\frac{1}{n}\Sigma''(x_i-\bar{x})^c\right\}^k \Big/ E\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^{\frac{1}{2}kc}, \tag{5·17}$$

the denominator of which is identical with the denominator of (4·1). Knowing the joint distribution (for normal samples) of $(x_1-\bar{x})$, $(x_2-\bar{x})$, ... (Geary, 1936), there is no theoretical difficulty in finding the mean values of the terms of the numerator for positive integer values of k. Here we shall be concerned only with the first and second moments, i.e. those for (5·17) for $k = 1$ and $k = 2$. We require the normal distribution of $z_1 = x_1-\bar{x}$ and the joint distribution of z_1 and $z_2 = x_2-\bar{x}$. These are

$$\begin{aligned}(z_1):&\quad \left(\frac{n}{2\pi\,\overline{n-1}}\right)^{\frac{1}{2}}\exp\left\{-\frac{nz_1^2}{2(n-1)}\right\}dz_1,\\ (z_1,z_2):&\quad \frac{1}{2\pi}\left(\frac{n}{n-2}\right)^{\frac{1}{2}}\exp\left\{-\frac{(n-1)\,(z_1^2+z_2^2)}{2(n-2)}-\frac{z_1z_2}{(n-2)}\right\}dz_1dz_2 = f(z_1,z_2)\,dz_1dz_2.\end{aligned} \tag{5·18}$$

Clearly the odd normal moments of $g(c)$ are zero. Then

$$E\left\{-\frac{1}{n}\Sigma'\,|\,x_i-\bar{x}\,|^c+\frac{1}{n}\Sigma''(x_i-\bar{x})^c\right\}^2 = \frac{n}{n^2}E\,|\,z_1\,|^{2c}+\frac{n(n-1)}{n^2}E_1(z_1,z_2), \tag{5·19}$$

where $E_1(z_1, z_2)$ is the mean value of the two-dimensional terms. We then have

$$\begin{aligned}E_1(z_1,z_2) &= \int_{-\infty}^{0}(-z_1)^c dz_1\int_{-\infty}^{0}(-z_2)^c dz_2 f(z_1,z_2)-\int_{-\infty}^{0}dz_1(-z_1)^c\int_0^{\infty}dz_2 z_2^c f(z_1,z_2)\\ &\quad -\int_0^{\infty}dz_1 z_1^c\int_{-\infty}^{0}dz_2(-z_2)^c f(z_1,z_2)+\int_0^{\infty}dz_1 z_1^c\int_0^{\infty}dz_2 z_2^c f(z_1,z_2)\\ &= \int_0^{\infty}\int_0^{\infty}z_1^c z_2^c dz_1 dz_2\{f(-z_1,-z_2)-f(-z_1,z_2)-f(z_1,-z_2)+f(z_1,z_2)\}\\ &= -\frac{2^{c+2}}{2\pi}\left(\frac{n}{n-2}\right)^{\frac{1}{2}}\frac{(n-2)^{c+1}}{(n-1)^{c+2}}\left(\frac{c}{2}\,!\right)^2\left\{1+\frac{(c+2)^2}{3!(n-1)^2}+\frac{(c+2)^2\,(c+4)^2}{5!(n-1)^4}+\dots\right\},\end{aligned} \tag{5·20}$$

$$Ez_1^{2c} = \frac{2c-1}{2}\,!\left(\frac{2\,\overline{n-1}}{n}\right)^c\frac{1}{\sqrt{\pi}}. \tag{5·21}$$

Also
$$E\left\{\frac{1}{n}\Sigma(x_i-\bar{x})^2\right\}^c = \left(\frac{2}{n}\right)^c\left(\frac{n+2c-3}{2}\right)!\Big/\left(\frac{n-3}{2}\right)!. \tag{5·22}$$

We now have all the expressions required for the variance of normal $g(c)$. We require, for what follows, only the term in n^{-1} which is

$$\sigma^2 = \frac{1}{n}\left\{\left(\frac{2c-1}{2}\right)!\,\frac{2^c}{\sqrt{\pi}}-\left(\frac{c}{2}\,!\right)^2\frac{2^{c+1}}{\pi}\right\}. \tag{5·23}$$

Consider now a field of alternative universes represented by

$$\frac{1}{\sqrt{(2\pi)}}\left\{1+\frac{\lambda_3}{6}(x^3-3x)\right\}e^{-\frac{1}{2}x^2}, \tag{5·24}$$

the 'first approximation to the law of error' (for universal variance unity), obviously the most appropriate asymmetrical field, for different values of the parameter λ_3, and con-

taining as a member of the field the normal distribution found for $\lambda_3 = 0$. For indefinitely large samples from (5·24) the mean value of $g(c)$ is

$$\delta = \frac{2\lambda_3}{6\sqrt{(2\pi)}}\int_0^\infty dx\, x^c(x^3-3x)\,e^{-\frac{1}{2}x^2} = \frac{c}{2}!\,(c-1)\,2^{\frac{1}{2}c}\frac{\lambda_3}{3}\bigg/\sqrt{\frac{1}{2\pi}}. \tag{5·25}$$

From (5·23) and (5·25)

$$\frac{\delta}{\sigma} = \frac{\lambda_3 n^{\frac{1}{2}}}{6}\tau(c), \tag{5·26}$$

the *skew discriminant* $\tau(c)$ being given by

$$\tau(c) = (c-1)\left\{\left(\frac{2c-1}{2}\right)!\left(\frac{c}{2}!\right)^{-2}\frac{\sqrt{\pi}}{2}-1\right\}^{-\frac{1}{2}} = (c-1)\left(\frac{2^{c+1}}{2c+1}\frac{I_{2c+2}}{I_{c+1}}-1\right)^{-\frac{1}{2}}. \tag{5·27}$$

Log-differentiating,

$$\frac{\tau'(c)}{\tau(c)} = \frac{1}{c-1} - \frac{2^{c+1}}{2}\left\{\frac{1}{2c+1}\left(\frac{2J_{2c+2}}{I_{c+1}} - \frac{I_{2c+2}J_{c+1}}{I^2_{c+1}}\right) - \frac{I_{2c+2}}{I_{c+1}}\frac{2}{(2c+1)^2} + \frac{I_{2c+2}}{I_{c+1}}\frac{\log 2}{2c+1}\right\}\left\{\frac{2^{c+1}}{2c+1}\frac{I_{2c+2}}{I_{c+1}}-1\right\}^{-1}. \tag{5·28}$$

Setting $c = 3$ and using (5·13) and (5·15), we find that $\tau'(3) = 0$. Values of $\tau(c)$ for four values of c are as follows:

c	$\tau(c)$	c	$\tau(c)$
2	2·370	4	2·389
3	2·450	5	2·236

Accordingly, for indefinitely large samples the test of asymmetry $g(c)$ is most efficient for $c = 3$, when the test becomes the familiar $\sqrt{b_1}$. The margin in favour of this value of c, as compared with others in the range $2 \leqslant c \leqslant 5$, is, however, quite small.

6. Tests of kurtosis from the power function viewpoint

It may be useful to open this section with an interpretation of the results of the previous section from the point of view of the type of error theory of J. Neyman & E. S. Pearson (1933, 1936). For this we consider two universes of the field, the normal W_0 and any non-normal universe W_1, and two tests of kurtosis $a(4) = b_2$ and $a(c_1)$ for a particular value c_1 of c. Suppose that samples are sufficiently large that $a(c)$, for samples from all universes of the field, may be regarded as normally distributed.

Given a probability α, a sample number n can be found so that the mean value of $a(c_1)$ from W_1 lies exactly at, say, the upper α probability point of the distribution of $a(c_1)$ from W_0. Then from the results established in the preceding section the value of $a(4)$ for the same sample of n from W_1 could lie beyond the α probability point of $a(4)$ for normal samples of n. Suppose that the rule adopted was to regard as non-normal all samples for which $a(c)$ lies beyond the normal α probability point, and suppose that a very large number N of samples were drawn, N_0 from universes not significantly different from normal (defining 'insignificance' in some manner) and N_1 from non-normal universes, so that $N = N_0 + N_1$, where N_0 and N_1 are not necessarily known in advance. Then using $a(c_1)$ the number of erroneous allocations will be approximately $\alpha N_0 + \frac{1}{2}N_1$, whereas using $a(4)$ the number will be $\alpha N_0 + (\frac{1}{2} - p)\,N_1$ ($\frac{1}{2} > p > 0$), showing a definite advantage in favour of $a(4)$. The same conclusion emerges whatever value of $c \neq 4$ or whatever non-normal universe be taken for comparison.

The type of error approach reveals the theoretical weakness of using the method of § 5 for the assessment of relative efficiency of tests of normality; namely that the proportion of

errors of judgment, even using $a(4)$, remains large, due fundamentally to concentrating on a single value (the mean) as typical or representative of samples from the non-normal universe; it is also a disadvantage that the sample number n_1 is necessarily a function of the particular value c_1 of c. The method has further disadvantages of which the principal are perhaps (i) a somewhat restricted field of alternative universes; (ii) the assumption that the samples were indefinitely large, essential to justify the normality of $a(c)$ for samples from any member of the universe field.

The Neyman-Pearson power function approach which will now be considered cannot be regarded as entirely free from these objections in its application to the material so far available from this research. It enables us, at any rate, to contemplate samples which, if not small, are within the range of experimental practicability.

The problem of the relative efficiency of the different members of a field of tests of kurtosis $a(c)$ will now be considered in its power function aspects. For the present purpose the *power* may be defined as follows:

Given a probability α (say 0·01), a sample number n, a particular value c_1 of c and a non-normal parent universe W_1, the power, in relation to these data, represents the frequency of $a(c_1)$ for samples drawn at random from W_1 lying beyond the α probability point for $a(c_1)$ computed from samples drawn from a normal universe. The greater the power the more discriminatory the test. Accordingly, it is in theory necessary to know the frequency distribution of $a(c)$ for all sample sizes, for all values of c and for all universes. Considering that the only frequency distribution of the field contemplated which can be regarded as determined for all sample sizes is $a(1)$ for normal samples (Geary, 1935, 1936), many compromises are necessary to give any kind of practical effect to the power concept. The compromises proposed are as follows:

(1) The form $a_1(c)$, given by (3·2), is used instead of the form $a(c)$ given by (3·1).

(2) Only large samples are dealt with.

(3) The field of alternative universes is restricted.

Using $a_1(c)$, the first four moments (from the origin) of $a_1(c)$ for samples from any universe can be expanded without real difficulty, and so approximate frequency distributions (using the Karl Pearson or Gram-Charlier systems) can be obtained. As to (1), from experiments in $a(1)$ and $a(4)$ the writer has verified that, for medium-sized normal samples, there is little difference between the probability points (e.g. 0·01, 0·05) of $a_1(c)$ and $a(c)$, though the higher semi-invariants (given n) are larger for the latter. In regard to (2) and (3) little confidence could be reposed in the values of the moments computed from expansions even to n^{-3} unless the sample number was at least of the order of 100 when c is greater than, say, 3; and, even if the moments were known exactly, the empirical frequencies would be more than doubtful for small samples. The approach finds its main justification in the consideration that any errors due to these necessary compromises may be presumed to apply more or less equally and in the same direction to the tests of kurtosis compared; generous, perhaps too generous, advantage is taken of this justification in the concluding part of this section.

Set, then,
$$a_1(c) = \left\{\frac{1}{n}\Sigma\,|\,x_i\,|^c\right\}\Big/\left\{\frac{1}{n}\Sigma x_i^2\right\}^{\frac{1}{2}c}, \tag{6·1}$$

so that
$$\frac{a_1(c)}{\alpha} = \left(1+\frac{1}{n}\Sigma y_i\right)\left(1+\frac{1}{n}\Sigma z_i\right)^{-\frac{1}{2}c}, \tag{6·2}$$

where
$$\alpha = \mu_{|c|}/\mu_2^{\frac{1}{2}c}, \quad y_i = (|\,x_i\,|^c - \mu_{|c|})/\mu_{|c|}, \quad z_i = (x_i^2 - \mu_2)/\mu_2, \tag{6·3}$$

the universal mean being taken as zero, without loss of generality. Raising (6·2) to powers 1, 2, 3, 4, expanding to the required degree the final factor, multiplying by the first factor on the right, and setting down the mean value of each term we find, to n^{-3},

$$\begin{aligned} M_1'/\alpha = 1 &- \frac{1}{n}\{k_1^{(1)}(11) - k_2^{(1)}(02)\} + \frac{1}{n^2}\{k_2^{(1)}(12) - k_3^{(1)}[(03) + 3(11)(02)] + 3k_4^{(1)}(02)^2\} \\ &+ \frac{1}{n^3}\{k_3^{(1)}[3(11)(02) - (13)] + k_4^{(1)}[(04) - 3(02)^2 + 4(11)(03) + 6(12)(02)] \\ &- k_5^{(1)}[10(03)(02) + 15(11)(02)^2] + 15k_6^{(1)}(02)\}, \end{aligned} \tag{6·4}$$

$$\begin{aligned} M_2'/\alpha^2 = 1 &+ \frac{1}{n}\{k_2^{(2)}(02) - 2k_1^{(2)}(11) + (20)\} + \frac{1}{n^2}\{-k_3^{(2)}(03) + 3k_4^{(2)}(02)^2 \\ &+ 2k_2^{(2)}(12) - 6k_3^{(2)}(11)(02) - k_1^{(2)}(21) + k_2^{(2)}(20)(02) + 2k_2^{(2)}(11)^2\} \\ &+ \frac{1}{n^3}\{k_4^{(2)}[(04) - 3(02)^2] - 10k_5^{(2)}(03)(02) + 15k_6^{(2)}(02)^3 - 2k_3^{(2)}[(13) - 3(11)(02)] \\ &+ 4k_4^{(2)}[2(11)(03) + 3(12)(02)] - 30k_5^{(2)}(11)(02)^2 \\ &+ k_2^{(2)}[(22) - (20)(02)] - k_3^{(2)}[(20)(03) + 3(21)(02)] - 2k_2^{(2)}(11)^2 \\ &- 6k_3^{(2)}(12)(11) + 12k_4^{(2)}(11)^2(02) + 3k_4^{(2)}(20)(02)^2\}, \end{aligned} \tag{6·5}$$

$$\begin{aligned} M_3'/\alpha^3 = 1 &+ \frac{1}{n}\{k_2^{(3)}(02) - 3k_1^{(3)}(11) + 3(20)\} + \frac{1}{n^2}\{-k_3^{(3)}(03) + 3k_4^{(3)}(02)^2 \\ &+ 3k_2^{(3)}(12) - 9k_3^{(3)}(11)(02) - 3k_1^{(3)}(21) + 3k_2^{(3)}(20)(02) + 6k_2^{(3)}(11)^2 \\ &+ (30) - 3k_1^{(3)}(20)(11)\} + \frac{1}{n^3}\{k_4^{(3)}[(04) - 3(02)^2] - 10k_5^{(3)}(03)(02) + 15k_6^{(3)}(02)^3 \\ &- 3k_3^{(3)}[(13) - 3(11)(02)] + 6k_4^{(3)}[2(11)(03) + 3(12)(02)] - 45k_5^{(3)}(11)(02)^2 \\ &+ 3k_2^{(3)}[(22) - (20)(02)] - 3k_3^{(3)}[(20)(03) + 3(21)(02)] + 9k_4^{(3)}(20)(02)^2 \\ &- 6k_2^{(3)}(11)^2 - 18k_3^{(3)}(12)(11) + 36k_4^{(3)}(11)^2(02) - k_1^{(3)}(31) \\ &+ k_2^{(3)}(30)(02) + 3k_1^{(3)}(20)(11) \\ &+ 3k_2^{(3)}[(20)(12) + 2(21)(11)] - 9k_3^{(3)}(20)(11)(02) - 6k_3^{(3)}(11)^3\}, \end{aligned} \tag{6·6}$$

$$\begin{aligned} M_4'/\alpha^4 = 1 &+ \frac{1}{n}\{k_2^{(4)}(02) - 4k_1^{(4)}(11) + 6(20)\} + \frac{1}{n^2}\{-k_3^{(4)}(03) + 3k_4^{(4)}(02)^2 \\ &+ 4k_2^{(4)}(12) - 12k_3^{(4)}(11)(02) - 6k_1^{(4)}(21) + 6k_2^{(4)}(20)(02) + 12k_2^{(4)}(11)^2 \\ &+ 4(30) - 12k_1^{(4)}(20)(11) + 3(20)^2\} + \frac{1}{n^3}\{k_4^{(4)}[(04) - 3(02)^2] \\ &- 10k_5^{(4)}(03)(02) - 15k_6^{(4)}(02)^3 - 4k_3^{(4)}[(13) - 3(11)(02)] \\ &+ 8k_4^{(4)}[2(11)(03) + 3(12)(02)] - 60k_5^{(4)}(11)(02)^2 + 6k_2^{(4)}[(22) - (20)(02)] \\ &- 6k_3^{(4)}[(20)(03) + 3(21)(02)] + 18k_4^{(4)}(20)(02)^2 - 12k_2^{(4)}(11)^2 \\ &- 36k_3^{(4)}(12)(11) + 72k_4^{(4)}(11)^2(02) - 4k_1^{(4)}(31) + 4k_2^{(4)}(30)(02) \\ &+ 12k_1^{(4)}(20)(11) + 12k_2^{(4)}[(12)(20) + 2(21)(11)] \\ &- 36k_3^{(4)}(20)(11)(02) - 24k_3^{(4)}(11)^3 + (40) - 4k_1^{(4)}(30)(11) - 3(20)^2 \\ &- 6k_1^{(4)}(20)(21) + 3k_2^{(4)}(20)^2(02) + 12k_2^{(4)}(20)(11)^2\}, \end{aligned} \tag{6·7}$$

where $$k_r^{(p)} = \frac{\frac{1}{2}pc(\frac{1}{2}pc+1)(\frac{1}{2}pc+2)\ldots(\frac{1}{2}pc+r-1)}{r!}, \quad (fg) = Ey_i^f z_i^g,$$

the latter, of course, the same for all i. The (fg) required for the computation of (6·4)–(6·7) are

$$\left.\begin{aligned}
(11) &= (\mu_{|2+c|}-\mu_2\mu_{|c|})/\mu_2\mu_{|c|},\\
(02) &= (\mu_4-\mu_2^2)/\mu_2^2,\\
(12) &= (\mu_{|4+c|}-2\mu_{|2+c|}\mu_2-\mu_{|c|}\mu_4+2\mu_{|c|}\mu_2^2)/\mu_{|c|}\mu_2^2,\\
(03) &= (\mu_6-3\mu_4\mu_2+2\mu_2^3)/\mu_2^3,\\
(04) &= (\mu_8-4\mu_6\mu_2+6\mu_4\mu_2^2-3\mu_2^4)/\mu_2^4,\\
(13) &= [\mu_{|6+c|}-3\mu_{|4+c|}\mu_2+3\mu_{|2+c|}\mu_2^2-\mu_{|c|}(\mu_6-3\mu_4\mu_2+3\mu_2^3)]/\mu_{|c|}\mu_2^3,\\
(21) &= [\mu_{|2c+2|}-2\mu_{|c+2|}\mu_{|c|}-\mu_2(\mu_{|2c|}-2\mu_{|c|}^2)]/\mu_{|c|}^2\mu_2,\\
(22) &= (\mu_{|2c+4|}-2\mu_{|2c+2|}\mu_2+\mu_{|2c|}\mu_2^2-2\mu_{|c+4|}\mu_{|c|}+4\mu_{|c+2|}\mu_{|c|}\mu_2-3\mu_{|c|}^2\mu_2^2+\mu_{|c|}^2\mu_4)/\mu_{|c|}^2\mu_2^2,\\
(20) &= (\mu_{|2c|}-\mu_{|c|}^2)/\mu_{|c|}^2,\\
(30) &= (\mu_{|3c|}-3\mu_{|2c|}\mu_{|c|}+2\mu_{|c|}^3)/\mu_{|c|}^3,\\
(31) &= (\mu_{|3c+2|}-3\mu_{|2c+2|}\mu_{|c|}+3\mu_{|c+2|}\mu_{|c|}^2-\mu_{|3c|}\mu_2+3\mu_{|2c|}\mu_{|c|}\mu_2-3\mu_{|c|}^3\mu_2)/\mu_{|c|}^3\mu_2,\\
(40) &= (\mu_{|4c|}-4\mu_{|3c|}\mu_{|c|}+6\mu_{|2c|}\mu_{|c|}^2-3\mu_{|c|}^4)/\mu_{|c|}^4.
\end{aligned}\right\} \qquad (6{\cdot}8)$$

(6·8) is, of course, an immediate consequence of (6·3). The writer has checked the accuracy of formulae (6·4)–(6·7) by reference to the normal universe for $c = 1$.

The reader will have no illusions as to the magnitude of the task of applying the foregoing theory to particular cases. The formulae are set down, however, in the hope that other researchers will be sufficiently sensible of the importance of the theory to assist in building up a fairly extensive set of results. The writer has to be content, in the meantime, to consider the case of the symmetrical universe field given by

$$\frac{1}{\sqrt{(2\pi)}}\left\{1+\frac{\lambda_4}{24}\left(\frac{d}{dx}\right)^4\right\}e^{-\frac{1}{2}x^2}, \qquad (6{\cdot}9)$$

when $\lambda_4 = \frac{1}{2}$, the normal being given, of course, for $\lambda_4 = 0$, and for $c = 4$ and $c = 1$. These values of c are selected because the theory in § 5 has suggested that $a(4)$ is probably the most efficient of the test-field $a(c)$, while $a(1)$ is the only member of the field for which the normal

Table 8. *Moments from formulae* (6·8)

(fg)	$c=4$		$c=1$	
	Normal	$\lambda_4=\frac{1}{2}$	Normal	$\lambda_4=\frac{1}{2}$
(11)	4	5·428571	1	1·17021276
(02)	2	2·5	2	2·5
(12)	24	45·64286	3	4·88297871
(03)	8	14	8	14
(04)	60	138	60	138
(13)	216	544·2857	21	44·106383
(21)	256/3	177·71428	1·141593	1·75544898
(22)	2,720/3	2,481·92857	7·707963	14·766814
(20)	32/3	16·142857	0·570796	0·63834981
(30)	352	799·142857	0·429204	0·6405182
(31)	4,352	12,785·2653	3	5·236134
(40)	23,552	73,250·178	—	2·002492

distribution is known for samples of all sizes. The necessary moments (fg) given by (6·8) are shown in Table 8. Based on the values in this table, moments (M') given by (6·4)–(6·7) of $a_1(c)$ and semi-invariants (L) derived therefrom are as follows. The normal values are, of course, known exactly but were computed for the purpose of checking the formulae:

$c = 4$; *normal universe*

$$\frac{L_1}{3} = \frac{M_1'}{3} \simeq 1 - \frac{2}{n} + \frac{4}{n^2} - \frac{8}{n^3},$$

$$\frac{M_2'}{9} \simeq 1 - \frac{4}{3n} - \frac{28}{n^2} + \frac{1040}{3n^3}, \quad \frac{L_2}{9} \simeq \frac{8}{3n} - \frac{40}{n^2} + \frac{1136}{3n^3},$$

$$\frac{M_3'}{27} \simeq 1 + \frac{2}{n} - \frac{48}{n^2} - \frac{1040}{n^3}, \quad \frac{L_3}{27} \simeq \frac{64}{n^2} - \frac{2368}{n^3},$$

$$\frac{M_4'}{81} \simeq 1 + \frac{8}{n} + \frac{40}{3n^2} - \frac{3520}{n^3}, \quad \frac{L_4}{81} \simeq \frac{3840}{n^3}.$$

$c = 4$; *universal* $\lambda_4 = \frac{1}{2}$

$$\frac{L_1}{3\cdot5} = \frac{M_1'}{3\cdot5} \simeq 1 - \frac{3\cdot357}{n} + \frac{11\cdot822}{n^2} + \frac{12\cdot1}{n^3},$$

$$\frac{M_2'}{(3\cdot5)^2} \simeq 1 - \frac{2\cdot286}{n} - \frac{57\cdot34}{n^2} + \frac{776\cdot03}{n^3}, \quad \frac{L_2}{(3\cdot5)^2} \simeq \frac{4\cdot4286}{n} - \frac{92\cdot25}{n^2} + \frac{831\cdot2}{n^3},$$

$$\frac{M_3'}{(3\cdot5)^3} \simeq 1 + \frac{3\cdot215}{n} - \frac{107\cdot47}{n^2} - \frac{2853\cdot89}{n^3}, \quad \frac{L_3}{(3\cdot5)^3} \simeq \frac{144\cdot61}{n^2} - \frac{6193\cdot95}{n^3},$$

$$\frac{M_4'}{(3\cdot5)^4} \simeq 1 + \frac{13\cdot143}{n} + \frac{20\cdot49}{n^2} - \frac{9529}{n^3}, \quad \frac{L_4}{(3\cdot5)^4} \simeq \frac{10{,}587}{n^3}.$$

$c = 1$; *normal universe*

$$L_1 = M_1' \simeq 0\cdot7978845608 + \frac{0\cdot19947114}{n} + \frac{0\cdot02493389}{n^2} - \frac{0\cdot03116737}{n^3},$$

$$L_2 \simeq \frac{0\cdot04507034}{n} - \frac{0\cdot07957747}{n^2} + \frac{0\cdot03978874}{n^3},$$

$$L_3 \simeq -\frac{0\cdot01685645}{n^2} + \frac{0\cdot07613597}{n^3}.$$

$c = 1$; *universal* $\lambda_4 = \frac{1}{2}$

$$\frac{L_1}{\mu_{|1|}} = \frac{M_1'}{\mu_{|1|}} \simeq 1 + \frac{0\cdot35239362}{n} - \frac{0\cdot159616}{n^2} - \frac{0\cdot745838}{n^3},$$

$$\frac{M_2'}{\mu_{|1|}^2} \simeq 1 + \frac{0\cdot79792429}{n} - \frac{0\cdot458012}{n^2} - \frac{1\cdot800648}{n^3}, \quad \frac{L_2}{\mu_{|1|}^2} \simeq \frac{0\cdot09313705}{n} - \frac{0\cdot262961}{n^2} - \frac{0\cdot196477}{n^3},$$

$$\frac{M_3'}{\mu_{|1|}^3} \simeq 1 + \frac{1\cdot336592}{n} - \frac{0\cdot850081}{n^2} - \frac{3\cdot239101}{n^3}, \quad \frac{L_3}{\mu_{|1|}^3} \simeq \frac{0\cdot053356}{n^2} + \frac{0\cdot204164}{n^3},$$

$$\mu_{|1|} = 0\cdot78126197.$$

Two sample sizes were considered: $n = 100$ and $n = 500$. For $n = 100$ and $c = 4$, the

following are the Pearson Type IV frequencies of $a_1(4)$ when the parent universes are normal and have $\lambda_4 = \beta_2 - 3 = \frac{1}{2}$ respectively:

$$\left.\begin{aligned} \text{Normal: } \lambda_4 = 0.\quad & \kappa \cos^{11\cdot3350}\theta\, e^{13\cdot015439}\,dx, \\ & \tan\theta = (x - 1\cdot873387)/0\cdot765849, \\ & \log_{10}\kappa = \bar{3}\cdot2644596. \end{aligned}\right\} \qquad (6\cdot10)$$

$$\left.\begin{aligned} \lambda_4 = \tfrac{1}{2}:\quad & \kappa \cos^{6\cdot0096}\theta\, e^{2\cdot3128\theta}\,dx, \\ & \tan\theta = (x - 2\cdot8522)/0\cdot9062, \\ & \log_{10}\kappa = \bar{1}\cdot7499974. \end{aligned}\right\} \qquad (6\cdot11)$$

The normal probability points shown in column (2) of Table 10 were derived from the foregoing normal frequency (6·10); the points in column (3) were derived from a Gram-Charlier formula (Geary, 1935). The 0·01 and 0·05 points given in column (2) are practically identical with those given by E. S. Pearson (1929) for $a(4)$, namely, 4·39 and 3·77. The powers given in column (4) are the aggregate frequencies lying beyond the values of the variate shown in column (2) on the assumption that the actual frequency was (6·11). The corresponding figures for $c = 1$ given in column (5) were based on a Gram-Charlier formula.

Table 9. *Power of* $a_1(c)$ *for* $c = 4$ *and* $c = 1$ *of discriminating* (6·9) *for* $\lambda_4 = \frac{1}{2}$ *from the normal* ($\lambda_4 = 0$) *at four normal theory probability levels. Samples of* 100

Normal theory probability	Normal theory probability points		Power for frequency (6·9) with $\lambda_4 = \frac{1}{2}$	
	$c = 4$ (upper)	$c = 1$ (lower)	$c = 4$	$c = 1$
(1)	(2)	(3)	(4)	(5)
0·01	4·3836	0·7482	0·0648	0·0695
0·05	3·7744	0·7642	0·1995	0·1979
0·10	3·5195	0·7725	0·3163	0·3037
0·20	3·3110	0·7824	0·4525	0·4597

Before discussing the comparative powers in Table 9 it will be convenient to give a table, 11, on the same lines but for $n = 500$. On account of the larger sample size it has been necessary to change the reference-probabilities given in column (1). For the construction of this table Gram-Charlier formulae were used throughout—the probability points being determined from the E. A. Cornish & R. A. Fisher (1937) formulae—after verifying that for two of the probability levels, 0·01 and 0·05, the probability points for $c = 4$ (column (2) above) did not differ appreciably from those given by E. S. Pearson, namely, 3·60 and 3·37 (for $a(4)$), based on a Type IV curve.

The analysis in § 5 has enabled us to come fairly firmly to the conclusion that for indefinitely large samples $a(4)$ was to be preferred to $a(1)$ as a test of normality. We see from Tables 9 and 10 that this is subject to an important qualification. Table 9 shows that the discriminating power is definitely greater for samples of 500 for $a(4)$ than for $a(1)$, but the superiority is less emphatic than might have been anticipated from § 5. For medium-sized samples (Table 9) $a(4)$ exhibits no superiority. Of course, these conclusions are very tentative, as being based upon a single alternative and on particular sample sizes. The writer had proposed, in addition, to examine the universes (i) $\lambda_3 = 0$, $\lambda_4 = 1$ and (ii) $\lambda_3^2 = \lambda_4 = \frac{1}{2}$ as alternatives to the normal but time did not permit; he ventures to repeat the hope that other students will take the matter up.

Table 10. *Power of $a_1(c)$ for $c = 4$ and $c = 1$ of discriminating (6·9) for $\lambda_4 = \frac{1}{2}$ from the normal ($\lambda_4 = 0$) at four probability levels. Samples of 500*

Normal probability	Normal probability points		Power for frequency (6·9) with $\lambda_4 = \frac{1}{2}$	
	$c=4$ (upper)	$c=1$ (lower)	$c=4$	$c=1$
(1)	(2)	(3)	(4)	(5)
0·005	3·7062	0·773167	0·1934	0·2067
0·01	3·6094	0·775684	0·2920	0·2790
0·05	3·3766	0·782482	0·5955	0·5196
0·10	3·2695	0·786058	0·7392	0·6509

7. CONCLUSION AND SUMMARY

In § 2 of the present paper it is shown that the actual probability of differences between means and variances derived from random samples on the nul-hypothesis may differ considerably from the probability derived from the standard tables (compiled on the assumption that the universal distribution is normal), when, in fact, the universal distribution is *not* normal. Accordingly, the standard tables cannot validly be used unless tests, based on the sample from which the inferences are to be drawn, or on a series of samples produced under similar conditions, have established the likelihood that the universal distribution is approximately normal. In certain cases—but these must be few—the nature of the material may, of itself, suffice to justify the assumption of universal normality. When universal normality cannot be assumed, the best course will be to correct the standard tables using, for this purpose, the moments (up to, say, the fourth) derived from the sample, in conjunction with the formulae given in § 2. This procedure is, of course, open to the objection that the moments derived from the sample may, in fact, differ substantially from the (in general unknown) universal moments, so that any probabilistic inference derived using sample moments must be accepted with reserve. If $b_2 = 3{\cdot}5$, say, it would be safer to assume that the universal value β_2 is 3·5, than to hope (without other evidence) that it is 3, the normal value; it might be 3·75 or even 4, when, usually, the standard table probabilities will be still further astray. It should not be difficult to construct supplementary tables giving very approximate corrections of the standard tables, using the moment expansions given in § 2, for different values of $\sqrt{\beta_1}$ and β_2. To compute unbiassed estimates of the latter, R. A. Fisher's k statistics (1929) should, of course, be used.

It may be asked if testing for normality and, when necessary, correction for universal non-normality is worth the trouble. To answer this question it is desirable to have regard to the logical position of the statistician, concerned with drawing inferences from samples, whose characteristic approach may be defined as *reductio ad paene absurdum*: if an event is highly improbable it must be regarded for practical purposes as impossible. St Thomas Aquinas's* famous 'certitude of probability' is peculiarly apt as applied to the mental attitude of the statistician, from two quite different viewpoints. The first is that decision, and action based on that decision, for which there is not certainty, but merely probabilistic preference, is absolute. One does not say that one has a preference of 20 to 1 for Fertilizer A

* 'According to the Philosopher, certitude is not to be sought equally in every matter.... Hence the certitude of probability suffices, such as may reach the truth in the greater number of cases, although it fails in the minority' (*Summa* IIa–IIae q. lxx, *a*. 2).

over Fertilizer B because the differences between the yields is at or near the 5 % probability point of some test functions: one necessarily decides without qualification that A is better than B.

The second aspect, which has the greater relevance in the present case, is that the statistician regards himself as endowed with 'certitude' when he knows that if he repeated an experiment, as to, say, significant differences in averages, a great number of times, he would be in error in attributing significant difference when, in fact, there was none, in a predetermined proportion of cases. He has certitude as to the probability though his decision in the individual case may be wrong. What is curious is that decisions (which, in effect, are absolute) can be based on probability levels which vary with the temperament of the statistician from perhaps a conservative 0·001 to a daring 0·1. For the particular statistician the probability level will vary with the case: for instance, the present writer would be inclined to suspect *non-normality near the 10 % probability level of the a(1) table, whereas he would not be* disposed to attach significance in, say, analysis of variance, until about the $2\frac{1}{2}$ % level. Naturally the level will depend on the importance attaching to the decision.

Since all the statistician usually requires from the table of probability for a given measure of significance is whether, on the nul-hypothesis, the probability is 'small', absolute precision is not necessary in the probability. If the probability is thought to be minute, say 0·001, it does not matter if in actual fact it is 0·002 or 0·0005. If, on the contrary, the standard table value is approaching the statistician's level of decision it surely matters a great deal: if he thinks his judgment is likely to be erroneous in 1 out of 20 experiments it must be of importance if, in fact, the true probability is something like 1 in 10 or 1 in 5. These are the kinds of contrasts that appear from § 2, from comparison of standard table probabilities with 'actual' probabilities found when the samples were assumed to be randomly drawn from certain arbitrarily selected types of non-normal universes. The computed probabilities in § 2 admittedly make no claim to exactitude in most of the cases, since the formulae were strained by their application to small sample theory. The point is, however, that the estimates of the actual probabilities are unbiassed in regard to the 'normal theory' probabilities: if the former could be closer to the latter, they might also be further away.

There is one case which is in a quite exceptional category, namely that considered at the beginning of § 2. As far as the writer is aware, this case has never been examined *theoretically* before, despite the extreme simplicity of the algebra. It is shown that in the simplest case of analysis of variance, when the two sample numbers are of the same order of magnitude, the variance is proportional, approximately, to (β_2-1), so that quite a small measure of universal kurtosis materially changes the probability. Statisticians must have been affected by a kind of hypnosis in favour of normal theory to have overlooked so trivial a point, a stricture from which the writer is not particularly concerned to exclude himself! An exception was E. S. Pearson (1931) who, on the basis of his results cited in § 2 (*a*), sounded a warning: 'The illustration should serve to emphasize the fact that certain of the "normal theory" tests can be used with greater confidence than others when dealing with samples from populations whose distribution laws are not known.'

An interesting chapter could be written on the fluctuations in the attitude of statisticians during the past century on the question of the occurrence of the normal frequency distribution in nature, a chapter, perhaps, in a large work on Fashions in the Sciences down the Ages. Amongst the following the historian may find the reasons for the prejudice in favour of the hypothesis of universal normality up to, say, the end of the last century:

(1) The fact that, to a close approximation, it applies in a wide range of *mathematical* conditions.

(2) The fact that the theory found practical applications predominantly in assessing the probability of errors in astronomical measurements and in games of chance where the mathematical model could reasonably be assumed to apply.

(3) The beauty of the mathematical theory and the facility of algebraic manipulation in the function involved.

(4) The general shape to the visual sense of such frequency distributions as were known, before χ^2 imposed its discipline.

With the development, about the beginning of the century, of the theory of moments, statisticians became almost over-conscious of universal non-normality. The concomitant semi-invariant approach had quite a different background. The difference between the moment and Karl Pearson curve system on the one hand and semi-invariants and the Gram-Charlier system on the other is fundamentally that for the former normality is a particular case like any other, whereas for the latter normality is basic and generative. Each system has its advantages and disadvantages as applied to the determination of frequency distributions of which the lower moments are known. In fanciful terms one might say that in the ship Gram-Charlier one might sail in perfect safety but only within limited, and more or less ascertainable, range of Port Normality, whereas in the good craft Pearson one can sail the seven seas—at one's own risk.*

Our historian will find a significant change of attitude about a quarter-century ago following on the brilliant work of R. A. Fisher who showed that, when universal normality could be assumed, inferences of the widest practical usefulness could be drawn from samples of any size. Prejudice in favour of normality returned in full force and interest in non-normality receded to the background (though one of the finest contributions to non-normal theory was made during the period by R. A. Fisher himself), and the importance of the underlying assumptions was almost forgotten. Even the few workers in the field (amongst them the present writer) seemed concerned to show that 'universal non-normality doesn't matter': we so wanted to find the theory as good as it was beautiful. References (when there were any at all) in the text-books to the basic assumptions were perfunctory in the extreme. Amends might be made in the interest of the new generation of students by printing in leaded type in future editions of existing text-books and in all new text-books:

Normality is a myth; there never was, and never will be, a normal distribution.

This is an over-statement from the practical point of view, but it represents a safer initial mental attitude than any in fashion during the past two decades.

As already indicated, the present work is incomplete, especially on the experimental side. The writer hopes that he has created a *prima facie* case for the importance of testing for normality.

SUMMARY

(i) Inferences drawn from the standard (normal) tables of z and t may be seriously in error if the conditions in which the standard tables apply (the principal of which is that the universes from which the samples are drawn are normal) are ignored.

* This comment must not be taken as applying to the problem of curve-fitting, i.e. to fitting a smooth curve to given frequencies, but to the problem of estimating the frequency function given the first few semi-invariants.

(ii) Sufficient conditions are given for the approach to normality, with increasing sample size, of the field of tests of normality $a(c)$ (given by (3·1)) for $c > 0$.

(iii) Many term expansions of the first four moments of $a(c)$ for normal samples are given with practical applications designed to find the values of c for which the moments could be used with confidence to find the frequency distributions for medium-size samples; semi-invariants of $a_1(2{\cdot}4)$ and $a_1(4)$ ($a_1(c)$ is given by (3·2)) are compared; correlations between $a_1(c)$ and $a_1(c')$ are examined.

(iv) For indefinitely large samples and a wide field of alternative universes $a(4)$ is found to be the most sensitive test of kurtosis and an analogous test of asymmetry $g(c)$ is found to be most sensitive for $c = 3$, $g(3)$ being the familiar $\sqrt{b_1}$.

(v) An examination of the relative efficiency of $a(1)$ and $a(4)$ from the Power Function point of view suggests that $a(4)$ is increasingly to be preferred as the sample size increases; for samples of moderate size $a(1)$ is probably as efficient as $a(4)$.

(vi) Throughout the paper a considerable range of formulae is given in case students may feel interested to carry the writer's researches a stage further so as to give a firmer basis to his conclusions or to modify them. It is suggested (§4) that the preparation of a table of probability points of $a(2{\cdot}2)$ for normal samples of different sizes be taken in hand.

REFERENCES

Baker, G. A. (1932). *Ann. Math. Statist.* **3**, 1.
Bartlett, M. S. (1935). *Proc. Camb. Phil. Soc.* **31**, 226.
Cornish, E. A. & Fisher, R. A. (1937). *Rev. Inst. Int. Statist.* **5**, 307.
Craig, C. C. (1928). *Metron*, **7**, 3.
Eden T. & Yates, F. (1933). *J. Agric. Sci.* **23**, 6.
Fisher, R. A. (1925). *Metron*, **5**, 90.
Fisher, R. A. (1929). *Proc. Lond. Math. Soc.* (2), **30**, 199.
Fisher, R. A. (1930). *Proc. Roy. Soc.* A, **130**, 16.
Fisher, R. A. & Wishart, J. (1931). *Proc. Lond. Math. Soc.* (2), **33**, 195.
Fréchet, M. (1937). *Géneralités sur les Probabilités. Variables aléatoires.*
Geary, R. C. (1933). *Biometrika*, **25**, 184.
Geary, R. C. (1935). *Biometrika*, **27**, 310, 353.
Geary, R. C. (1936). *J. Roy. Statist. Soc.* (*Supplement*), **3**, 178.
Geary, R. C. (1936). *Biometrika*, **28**, 295.
Geary, R. C. (1947). *Biometrika*, **34**, 68.
Geary, R. C. & Pearson, E. S. (1938). *Tests of Normality.*
Geary, R. C. & Worlledge, J. P. G. (1946). *Biometrika*, **34**, 98.
Gosset, W. S. (1908). *Biometrika*, **6**, 1.
Hsu, C. T. & Lawley, D. N. (1940). *Biometrika*, **31**, 238.
Kendall, M. G. (1941). *Biometrika*, **32**, 81.
Nair, A. N. K. (1942). *Sankhyā*, **5**, 393.
Neyman, J. & Pearson, E. S. (1933). *Philos. Trans.* A, **231**, 289.
Neyman, J. & Pearson, E. S. (1936). *Statist. Res. Mem.* **1**, 1.
Pearson, E. S. (1929). *Biometrika*, **21**, 337.
Pearson, E. S. (1930). *Biometrika*, **22**, 239.
Pearson, E. S. (1931*a*). *Biometrika*, **22**, 423.
Pearson, E. S. (1931*b*). *Biometrika*, **23**, 114.
Pearson, E. S. (1935). *Biometrika*, **27**, 333.
Pearson, E. S. & Adyanthaya, N. K. (1929). *Biometrika*, **21**, 259.
Pearson, Karl (1895). *Philos. Trans.* A, **186**, 343.
Pepper, Joseph (1932). *Biometrika*, **24**, 55.
Rider, P. R. (1931). *Ann. Math. Statist.* **2**, 48.
Rietz, H. L. (1939). *Ann. Math. Statist.* **10**, 265.
Shewhart, W. A. & Winters, F. W. (1928). *J. Amer. Statist. Ass.* **23**, 144.
Wishart, J. (1930). *Biometrika*, **22**, 224.
Yasukawa, K. (1934). *Tokohu Math. J.* **38**, 465.

THE STRATIFIED SEMI-STATIONARY POPULATION

By S. VAJDA

1. Constant population

Let a set of non-increasing real values $p_0 = 1, p_1, \dots, p_n, p_{n+1} = 0$ be given, and let p_i represent the probability of a person of age 0 surviving the i following years. Further, let $l_0, l_1, \dots, l_n$ represent the numbers of persons of age $0, 1, \dots, n$ living at time $t = 0$. We consider then the development of such a population during the years following $t = 0$, under the assumption that the probabilities p_i remain the same throughout the period investigated.

Only persons of age 0 are to enter the population, and the number of such entrants shall be such that the total of the population is kept constant at a number $H = \sum_{i=0}^{n} l_i$. At the end of the first year the survivors of the H persons who were alive at $t = 0$ will be (if we put $l_i/p_i = r_i$, say)

$$\sum_{i=0}^{n-1} l_i \frac{p_{i+1}}{p_i} = \sum_{i=0}^{n-1} r_i p_{i+1} \leqslant H,$$

and therefore the number of entrants at the beginning of the second year (i.e. at $t = 1$) is $\phi_1 = H - \sum_{i=0}^{n-1} r_i p_{i+1}$.

By the same argument the entrants at $t = 2$ will be

$$\phi_2 = H - \phi_1 p_1 - \sum_{i=0}^{n-2} r_i p_{i+2},$$

and so on; generally

$$\phi_t = H - \phi_{t-1} p_1 - \phi_{t-2} p_2 - \dots - \phi_1 p_{t-1} - \sum_{i=0}^{n-t} r_i p_{i+t}, \tag{1}$$

as long as $t \leqslant n$, that is, as long as there are survivors of the initial population. For $t > n$ we obtain

$$H = \phi_t + \phi_{t-1} p_1 + \phi_{t-2} p_2 + \dots + \phi_{t-n} p_n. \tag{2}$$

We want to find an expression for ϕ_t, which must obviously depend on $l_1, l_2, \dots, l_n$. Now (2) is a difference equation for the function ϕ_t of t and can easily be solved. For this purpose consider the 'characteristic equation'

$$x^n + x^{n-1} p_1 + x^{n-2} p_2 + \dots + x p_{n-1} + p_n = 0. \tag{3}$$

Let this equation have the roots $x_1, x_2, \dots, x_r$, where x_i is a k_i-fold root and $x_i \neq x_j$. We have then as a solution of the difference equation (2)

$$\phi_t = H_1 + P_1(t)\, x_1^t + \dots + P_r(t)\, x_r^t, \tag{4}$$

where $H_1 = H/\Sigma p_i$ and $P_i(t) = \alpha_{i1} + \alpha_{i2} t + \dots + \alpha_{ik_i} t^{k_i - 1}$. The α_{ij} must be found from the initial population, i.e. from equations of the form (1) which contain the first n numbers of entrants $\phi_1, \phi_2, \dots, \phi_n$. But we find by inspecting these equations, which are of the form ($t \leqslant n$)

$$H = \phi_t + \phi_{t-1} p_1 + \dots + \phi_1 p_{t-1} + r_0 p_t + r_1 p_{t+1} + \dots + r_{n-t} p_n,$$

that they are equivalent to

$$r_t = \phi_{-t} = H_1 + \sum_{i=1}^{r} \sum_{j=1}^{k_i} \alpha_{ij} x_i^{-t} (-t)^{j-1}. \tag{5}$$

Hence the α_{ij} can be fixed, dependent on the $r_i = l_i/p_i$ and thus on the initial population.

We have thus proved:

If a population with an age distribution $l_0, l_1, \ldots, l_n$ is subject to survival rates p_i ($i = 1, 2, \ldots, n$), and if this population is kept constant by ϕ_t entrants of age 0 at the end of the tth year, then ϕ_t is given by (4), *where the x_i are the different roots of* (3), *and the α_{ij} must be found from the set* (5).

The population after t years will have the following age distribution:

$$\phi_t, \quad \phi_{t-1}p_1, \quad \phi_{t-2}p_2, \quad \ldots, \quad \phi_{t-n}p_n.$$

It can easily be proved that, if the p_i are decreasing (and not merely non-increasing), then for all the roots x_i of equation (3) we have $|x_i| < 1$ and that any real root must be negative. Hence the ϕ_t will oscillate around their limit $\lim_{t=\infty} \phi_t = H_1$. The age distribution of the population thus tends, again through oscillations, to $H_1, H_1p_1, \ldots, H_1p_n$, which may be called the *intrinsic stationary population.* Obviously, if the initial population has already this distribution, it will not alter any more and the number of entrants will be constant and $= H_1$. In such a case all $\alpha_{ij} = 0$, and $r_i = H_1 = l_0$, whatever the x_i may be.

On the other hand, if $p_i = p_{i+1}$ holds for one or more values of i, then we may get cycles, and this is easily seen for the equation $x^n + x^{n-1} + \ldots + x + 1 = 0$. All roots have modulus 1, and it depends on the initial population whether we are dealing with the stationary case or with periodic cycles. No tendency towards an intrinsic stationary population appears in such a case.

Example. Let us assume that we have the following probabilities of survival:

p_1	p_2	p_3	p_4	p_5	p_6
7/8	49/96	5/32	13/384	1/384	0

The characteristic equation can then be written

$$384x^5 + 336x^4 + 196x^3 + 60x^2 + 13x + 1 = 0,$$

which has the five different roots

$$-\tfrac{1}{4} \pm \sqrt{-\tfrac{5}{48}}, \quad -\tfrac{1}{8} \pm \sqrt{-\tfrac{7}{64}} \quad \text{and} \quad -\tfrac{1}{8}.$$

The initial population will be assumed to be

l_0	l_1	l_2	l_3	l_4	l_5
859	1269	229	50	115	56

which implies $H_1 = 1000$ (approx.). Therefore the $r_i = l_i/p_i$ are

r_0	r_1	r_2	r_3	r_4	r_5
859	1450·38	448·86	319·14	3405·42	21420·90

From $r_k = 1000 + \alpha_1 x_1^{-k} + \alpha_2 x_2^{-k} + \ldots + \alpha_5 x_5^{-k}$ ($k = 1, 2, \ldots, 5$) we find

α_1	α_2	α_3	α_4	α_5
$-70+60i$	$-70-60i$	0	0	-1

It follows that the number of entrants in the year t ($= 0, 1, 2, \ldots$) will be

$$\phi_t = 1000\left[1 + (-0{\cdot}07 + 0{\cdot}06i)\left(-\tfrac{1}{4} + \sqrt{-\tfrac{5}{48}}\right)^t + (-0{\cdot}07 - 0{\cdot}06i)\left(-\tfrac{1}{4} - \sqrt{-\tfrac{5}{48}}\right) + \frac{-0{\cdot}001}{(-8)^t}\right].$$

These numbers are given in the first row of Table 1, which shows the evolution of the whole population.

Table 1

t	0	1	2	3	4	5	6	7	8 and after
Age 0	859	996	1025	989	1001	1001	999	1000	1000
1	1269	752	872	897	865	876	876	874	875 ($=1000\times 7/8$)
2	229	740	438	508	523	505	511	511	510 ($=1000\times 49/96$)
3	50	70	227	134	156	160	155	156	156 ($=1000\times 5/32$)
4	115	11	15	49	29	34	34	34	34 ($=1000\times 13/384$)
5	56	9	1	1	4	2	3	3	3 ($=1000\times 1/384$)
	2578	2578	2578	2578	2578	2578	2578	2578	2578

2. Two constant populations

All this covers well-known ground.* A new problem arises, however, when we consider two initial populations with two sets of probabilities of survival, say p_i $(i = 1, 2, \ldots, n_1)$ and $\bar{p}_i$ $(i = 1, 2, \ldots, n_2)$, where $p_0 = \bar{p}_0 = 1$ and $p_{n_1} . p_{n_2} \neq 0$. We ask now whether it is possible to keep both constant by the same number of yearly entrants. More precisely:

Let the two equations

$$\sum_{i=0}^{n_1} p_i x^{n_1-i} = 0 \quad \text{and} \quad \sum_{i=0}^{n_2} \bar{p}_i y^{n_2-1} = 0$$

have the roots $x_1, \ldots, x_r$ with multiplicities $k_1, \ldots, k_r$ and $y_1, \ldots, y_s$ with multiplicities $j_1, \ldots, j_s$ respectively. No two x_i or two y_i are equal and no x_i or y_i is zero. Under what further conditions, concerning the x's and the y's, can the expressions ϕ_t and ψ_t then have the same numerical values for all integral values of t, i.e.

$$(H_1 - \bar{H}_1) + \sum_{i=1}^{r} P_i(t)\, x_i^t - \sum_{i=1}^{s} \bar{P}_i(t)\, y_i^t = 0 \quad \text{for} \quad t = 0, 1, 2, \ldots, \tag{6}$$

where $$\phi_t = H_1 + \sum_{i=1}^{r} P_i(t)\, x_i^t \quad \text{with} \quad P_i(t) = \alpha_{i1} + \alpha_{i2} t + \ldots + \alpha_{ik_i} t^{k_i-1}$$

and $$\psi_t = \bar{H}_1 + \sum_{i=1}^{s} \bar{P}_i(t)\, y_i^t \quad \text{with} \quad \bar{P}_i(t) = \beta_{i1} + \beta_{i2} t + \ldots + \beta_{ij_i} t^{j_i-1}?$$

Suppose first that none of the x's equals any of the y's. Then it is known that the determinant of any set of equations of the system (6) is not zero. It follows that we must have $H_1 = \bar{H}_1$ and all α's and β's $= 0$, hence all $P_i(t)$ and $\bar{P}_i(t) \equiv 0$. In this case the two populations must already be stationary and therefore identical with the intrinsic stationary populations which are implied by the sets p_i and $\bar{p}_i$, respectively.

On the other hand, if some of the x's are equal to some of the y's, say $x_1 = y_1, \ldots, x_m = y_m$ and all the others are different, then we find by the same argument that $H_1 = \bar{H}_1$ and $P_i = \bar{P}_i$ for the first m values of i, whereas all the other P_i and $\bar{P}_i$ are identically zero. (It is, of course, again possible that *all* the P_i and $\bar{P}_i$ are identically zero and that we have, in fact, again the two intrinsic stationary populations.)

If all x's are equal to the y's, with equal multiplicities, then the two equations are equal

* It follows, for example, from results of P. H. Leslie (1945).

and the two populations must be identical, if they are to be kept constant by equal numbers of entrants.

We have thus reached the following conclusion: If we assume that the two equations given above are not identical, and that the initial populations are not the intrinsic stationary ones, then they must be such that the two equations have some (but not all) roots equal and if we calculate the corresponding P_i and $\bar{P}_i$ (see (4) and (5) of the previous section), then those corresponding to the equal roots must be identical and the others must vanish. This includes the case where the x's and y's are the same, but with different multiplicities, so that the P_i and $\bar{P}_i$ do not all extend to the highest power of t which would be admissible by (7) or (8) respectively.

If the two populations are the intrinsic stationary ones, then the numbers of entrants will be constant (i.e. independent of the year) and the two constants will be equal if and only if

$$\frac{\Sigma l_i}{\Sigma p_i} = \frac{\Sigma \bar{l}_i}{\Sigma \bar{p}_i},$$

where the first expression refers to the first and the second expression to the second population.

Example. In the example used in § 1 we have $\alpha_3 = \alpha_4 = 0$, and we can therefore try to obtain a second population which is kept constant by the same numbers of entrants as the first one. We construct an equation which has again the roots $-\frac{1}{4} \pm \sqrt{-\frac{5}{48}}$ and $-\frac{1}{8}$, but not $-\frac{1}{8} \pm \sqrt{-\frac{7}{64}}$. Such an equation is, for example,

$$480x^5 + 396x^4 + 218x^3 + 62x^2 + 13x + 1 = 0,$$

which has the roots $-\frac{1}{4} \pm \sqrt{-\frac{5}{48}}$, $-\frac{1}{8}$, and $-\frac{1}{10} \pm \sqrt{-\frac{9}{100}}$. It implies that the probabilities of survival, i.e. $\bar{p}_i$, are

$$33/40 \quad 109/240 \quad 31/240 \quad 13/480 \quad 1/480,$$

and as the ϕ_t (and the $r_t = \phi_{-t}$) are to be the same as in § 1, the initial population must now be $\bar{l}_i = r_i \bar{p}_i$. Table 2 shows the development of such a population, and it will be seen that the first line is identical with that in Table 1.

Table 2

t	0	1	2	3	4	5	6	7	8 and after
Age 0	859	996	1025	989	1001	1001	999	1000	1000
1	1196	709	822	845	816	826	826	824	825 (= 1000 × 33/40)
2	204	658	390	452	465	449	455	455	454 (= 1000 × 109/240)
3	41	58	187	111	129	132	128	129	129 (= 1000 × 31/240)
4	92	9	12	39	23	27	27	27	27 (= 1000 × 13/480)
5	45	7	1	1	3	2	2	2	2 (= 1000 × 1/480)
	2437	2437	2437	2437	2437	2437	2437	2437	2437

3. Stratified population: two grades

The results of the previous sections will now be used for an investigation of the stratified population.* First, we consider a population split into a lower and a higher grade in the following way:

We assume that all members of age 0 are in the lower grade only, but that all other ages may share in both grades. Apart from mortality, which operates on all members according to their age, we assume that at every age a certain proportion dependent on that age is 'promoted', at the end of the year, from the lower into the higher grade. Our problem is to discover whether this can be done whilst maintaining the totals in both grades constant; naturally the grand total of the population must remain constant.

It is sufficient to deal only with the lower grade, as the numbers at each age in the higher one can be found by subtracting those in the lower grade from the total population at that age. Now the lower grade is depleted by mortality and also by promotions. If the probability of remaining unpromoted until age i is t_i, then the probability of not leaving the grade in this period is $p_i t_i = \bar{p}_i$, say. Since all entrants into the population are at the same time entrants into the lower grade, our problem thus reduces to the following:

Is it possible to find an initial population, stratified into two grades, such that, on the basis of mortality described by p_i, the number of entrants every year necessary to keep the population constant is the same as that calculated on the basis of mortality-cum-promotion, described by $\bar{p}_i$?

We can apply our results in § 2 to this case by considering the lower grade and the total population as the two populations given. It follows that the lower grade can only be kept constant by that number of entrants which is necessary for the total population, if the latter is initially such that some of the $P_i(t)$ which depend on it are either identically zero or at least do not extend to the highest degree indicated by the multiplicities of the corresponding roots in $\Sigma p_i x^{n-i} = 0$. In order to find a suitable initial population for the lower grade it is then necessary to find an equation $\Sigma \bar{p}_i y^{n-i} = 0$ which has the roots, with the necessary multiplicities, which appear explicitly in ϕ_t as calculated from the original equation, but which is not identical with it. The degree of $\Sigma \bar{p}_i y^{n-i} = 0$ may be lower than or equal to that of $\Sigma p_i x^{n-i} = 0$. If it is lower, then all members of the population will be in the higher grade at the highest age or ages.

This condition is not sufficient, however. In view of the interpretation of the equation containing the $\bar{p}_i$'s these coefficients must be positive and, as the lower grade is a part of the whole, we must have $\bar{p}_i \leqslant p_i$ for all i. But it is not necessary that we have also $\bar{p}_{i+1} \leqslant \bar{p}_i$. If the opposite holds, this could still bear a practical interpretation. It would mean that reversions occur from the higher into the lower grade.

If an equation with the necessary and sufficient properties can be found, then we take the $r_i = l_i/p_i$ which we had to start with and construct the initial population of the lower grade by writing the number at age i as $\bar{l}_i = r_i \bar{p}_i = l_i \bar{p}_i / p_i$.

It will be seen that in such a population the age distributions change with the passage of time (tending to a stationary limit) but that nevertheless all entrants have the same combined prospects of survival and promotion. (Thus from the point of view of a member of the community his position is the same as if he entered a stationary population. His chances

* *Cf., for the stationary case, with continuous changes,* H. L. Seal (1945).

of promotion are unaffected by the changes in the age distribution of those in front of him. But the characteristics of the population as a whole, for instance the efficiency of the staff from the point of view of an employer may, of course, vary considerably.) Such a population will be called *semi-stationary*.

Example. The population shown in Table 2 can be taken as representing a lower grade within the population given in Table 1. The ratios $t_i = \overline{p}_i/p_i$ are then:

t_0	t_1	t_2	t_3	t_4	t_5
1	33/35	218/245	62/75	4/5	4/5

Table 3 is constructed by subtracting Table 2 from Table 1 and thus shows the composition of the higher grade.

Table 3

t	0	1	2	3	4	5	6	7	8 and after
Age 1	73	43	50	52	49	50	50	50	50
2	25	82	48	56	58	56	56	56	56
3	9	12	40	23	27	28	27	27	27
4	23	2	3	10	6	7	7	7	7
5	11	2	—	—	1	—	1	1	1
	141	141	141	141	141	141	141	141	141

4. Stratified population: more than two grades

Let us now split up the higher grade as well. We have then, say, k grades, with grades 2 and above forming the aggregate which was simply called the higher grade in § 2; grade 1 is identical with the lower grade of that section.

We assume further that promotions from any grade into the next higher one take place at the end of every year and that every promotee into any grade has to stay there for at least one year. Thus in any population the lowest possible age of grade g is $g-1$. The actual lowest ages may be different, because the first promotion rates different from 0 may concern higher ages than these. The rates of promotion can be different from grade to grade, but depend within each grade only on the age, as before.

We shall again investigate whether it is possible to keep the total numbers of every grade constant, even if the age distributions of the grades are changing.

We have seen that the age distribution of the total population, after t years, is

$$\phi_t, \quad \phi_{t-1}p_1, \quad \ldots, \quad \phi_{t-n}p_n.$$

The distribution of grade 1 is, at the same time,

$$\phi_t, \quad \phi_{t-1}\overline{p}_1, \quad \ldots, \quad \phi_{t-n}\overline{p}_n,$$

and it is assumed that the set of p_i is not identical with the set of $\overline{p}_i$. Hence grades 2 and above will have the age distribution

$$\phi_t(p_0-\overline{p}_0), \quad \phi_{t-1}(p_1-\overline{p}_1), \quad \phi_{t-n}(p_n-\overline{p}_n).$$

Let us assume that $\phi_{t-\nu}(p_\nu-\overline{p}_\nu)$ is the first item in this series which is not zero. Clearly we have $\nu \geqslant 1$. Then, as far as numbers of members (and not their individual careers) are con-

cerned, this aggregate of grades 2 and above is equivalent to a population which has arisen from successive annual entrants $\phi_{t-\nu}(p_\nu-\overline{p}_\nu)$ who have been subject to rates of survival

$$q_1=\frac{p_{\nu+1}-\overline{p}_{\nu+1}}{p_\nu-\overline{p}_\nu},\quad \ldots,\quad q_{n-\nu}=\frac{p_n-\overline{p}_n}{p_\nu-\overline{p}_\nu}.$$

It must be understood, however, that 'survival' is here a balance between deaths and promotions into the grade, so that these rates may very well exceed unity.

The number of annual entrants into grade 2 is given by

$$\phi_{t-\nu}(p_\nu-\overline{p}_\nu)=\left[H_1+\sum_{i=1}^{m}P_i(t-\nu)\,x_i^{t-\nu}\right](p_\nu-\overline{p}_\nu),$$

where $x_1,\ldots,x_m$ are the common roots of $\sum_{i=0}^{n}p_ix^{n-i}=0$ and $\sum_{i=1}^{n}\overline{p}_iy^{n-i}=0$, with multiplicities k_i and j_i respectively, and where the P_1 are polynomials whose order does not exceed either k_i-1 or j_i-1. (They may all be identically zero.)

The x_i are, of course, also roots of $\sum_{i=\nu}^{n}(p_i-\overline{p}_i)\,x^{n-i}=0$, with multiplicities given by the smaller of k_i and j_i.

We ask now if it is possible to construct grade 2 alone in such a way that its total remains also constant. The argument which has been used in §3 shows that this is possible if another equation of grade $n-\nu$ can be found whose coefficients w_i, say, are not larger than the corresponding q_i (and $w_0=1$), which has once again the roots $x_1,\ldots,x_m$, with multiplicities g_i at least. If $g_1+\ldots+g_m=n-\nu$, then this is clearly impossible. If $g_1+\ldots+g_m$ is smaller than this value, then we can try to find such an equation. The initial population can also be then found, if we multiply the initial population of grades 2 and above by w_i/q_i. The grades 1, 2 and the aggregate of 3 and above can then be constructed and every stratum kept constant, but with changing age distributions.

We can proceed in the same way and find at each step whether further splitting up is possible beyond 3 grades, 4 grades, etc. It is seen that in general, if $\Sigma g_i=n-m$, and if grade g starts in fact at age $g-1$, then $m+1$ grades can exist.

The smallest value of Σg_i is 1, and in this extreme case n grades can be constructed, i.e. one less than the number of ages. The nth grade will then contain the ages $n-1$ and n. Further, since x_1 is a root of $x-x_1=0$, the age distribution of this highest grade is

$$H_1+\alpha_1x_1^{t-n+1},\quad -(H_1+\alpha_1x_1^{t-n})\,x_1=-H_1x_1-\alpha_1x_1^{t-n+1}$$

(x_1 is, of course, negative).

Example. We use again the same example as before. The characteristic equation for the whole population was

$$x^5+\tfrac{7}{8}x^4+\tfrac{49}{96}x^3+\tfrac{5}{32}x^2+\tfrac{13}{384}x+\tfrac{1}{384}=0,$$

and that for grade 1 alone

$$x^5+\tfrac{33}{40}x^4+\tfrac{109}{240}x^3+\tfrac{31}{240}x^2+\tfrac{13}{480}x+\tfrac{1}{480}=0.$$

The difference between these two equations gives the equation for grade 2 and above

$$x^4+\tfrac{9}{8}x^3+\tfrac{13}{24}x^2+\tfrac{13}{96}x+\tfrac{1}{96}=0.$$

This equation has, of course, the roots $-\frac{1}{4}\pm\sqrt{-\frac{5}{48}}$, $-\frac{1}{8}$ which are common to the two characteristic equations of the fifth degree, and also a further root $-\frac{1}{2}$. Now there is

a biquadratic equation with the three specified common roots and not larger coefficients (and having the coefficient of x^4 equal to unity), viz.

$$x^4+\tfrac{33}{40}x^3+\tfrac{17}{48}x^2+\tfrac{1}{15}x+\tfrac{1}{240}=0.$$

The fourth, irrelevant, root is $-1/5$. This equation leads to the following development:

Grade 2 only

t	0	1	2	3	4	5	6	7	8 and after
Age 1	73	43	50	52	49	50	50	50	50
2	18	60	35	41	43	40	41	41	41
3	6	8	26	14	18	19	18	18	18
4	11	—	1	5	2	3	3	3	3
5	4	1	—	—	—	—	—	—	—
	112	112	112	112	112	112	112	112	112

Grade 3 and above

Age 2	7	22	13	15	15	16	15	15	15
3	3	4	14	9	9	9	9	9	9
4	12	2	2	5	4	4	4	4	4
5	7	1	—	—	1	—	1	1	1
	29	29	29	29	29	29	29	29	29

Analysis into further grades is impossible in this case, because the characteristic equation of the third grade does not have any roots apart from the three common roots of all previous equations.

5. Promotion rates dependent on seniority

We still consider more than two grades, but now we will assume that the promotion rates do not depend on the attained age but on the seniority, i.e. on the time spent in the grade, instead. In the lowest grade seniority is equivalent to age, because all members were supposed to enter at the lowest age only. If we consider again the two grades of § 3, but this time take note of differences in seniority, we find the following pattern:

Age	Lower grade	Higher grade				Total
		Seniority				
		0	1		$x-1$	
0	ϕ_t					ϕ_t
1	$\phi_{t-1}p_1t_1$	$\phi_{t-1}p_1(t_0-t_1)$				$\phi_{t-1}p_1$
2	$\phi_{t-2}p_2t_2$	$\phi_{t-2}p_2(t_1-t_2)$	$\phi_{t-2}p_2(t_0-t_1)$			$\phi_{t-2}p_2$
...	....	....	....	...	....	....
x	$\phi_{t-x}p_xt_x$	$\phi_{t-x}p_x(t_{x-1}-t_x)$	$\phi_{t-x}p_x(t_{x-2}-t_{x-1})$	...	$\phi_{t-x}p_x(t_0-t_1)$	$\phi_{t-x}p_x$
...	....	....	....	...	....	....

Note. $t_i=\bar{p}_i/p_i$ and hence $t_0=1$.

If we consider now promotion from grade 2 into grade 3, and if we introduce u_s, the probability of not being promoted during s years from grade 2 ($u_0 = 1$), we see that grades 2 and 3 (including higher grades, if any) will have the following constitution:

Grade 2

Age	Seniority 0	1		$x-1$
1	$\phi_{t-1}p_1(t_0-t_1)\,u_0$			
2	$\phi_{t-2}p_2(t_1-t_2)\,u_0$	$\phi_{t-2}p_2(t_0-t_1)\,u_1$		
...	....	....	...	....
x	$\phi_{t-x}p_x(t_{x-1}-t_x)\,u_0$	$\phi_{t-x}p_x(t_{x-2}-t_{x-1})\,u_1$	...	$\phi_{t-x}p_x(t_0-t_1)\,u_{x-1}$
...	....	....	...	....

Grade 3

Age	Seniority 0	1		$x-2$
2	$\phi_{t-2}p_2(t_0-t_1)\,(u_0-u_1)$			
3	$\phi_{t-3}p_3[(t_1-t_2)\,(u_0-u_1)$ $+(t_0-t_1)\,(u_1-u_2)]$	$\phi_{t-3}p_3(t_0-t_1)\,(u_0-u_1)$		
...	....	....	...	....
x	$\phi_{t-x}p_x[(t_{x-2}-t_{x-1})\,(u_0-u_1)$ $+(t_{x-3}-t_{x-2})\,(u_1-u_2)+\dots$ $+(t_0-t_1)\,(u_{x-2}-u_{x-1})]$	$\phi_{t-x}p_x[(t_{x-3}-t_{x-2})\,(u_0-u_1)$ $+(t_0-t_1)\,(u_{x-3}-u_{x-2})]$	...	$\phi_{t-x}p_x(t_0-t_1)\,(u_0-u_1)$
...	....	....	...	....

It follows by means of the same argument as before that grade 2 can be kept constant if we can find the u_i such that the equation

$$x^{n-1}p_1(t_0-t_1)+x^{n-2}p_2[(t_1-t_2)+(t_0-t_1)\,u_1]+\dots$$
$$+p_n[(t_{n-1}-t_n)+(t_{n-2}-t_{n-1})\,u_1+\dots+(t_0-t_1)\,u_{n-1}] = 0$$

has the same roots which were common to $x^n+x^{n-1}p_1+\dots+p_n = 0$ and

$$x^{n-1}p_1(t_0-t_1)+x^{n-2}p_2(t_0-t_2)+\dots+p_n(t_0-t_n) = 0,$$

which is identical with the difference of the first two equations of degree n, referring respectively to the whole population and to the lowest grade. We must further insist that all u_i must have non-negative values, not larger than 1. The coefficients of the powers of x must also be positive, but it is not necessary that $u_{i+1} \leqslant u_i$, unless we do not admit reversions. If m is the number of common roots, then it follows again as in the last section that $n-m+1$ grades could exist which remain constant under the operation of promotions, but that their age and seniority distributions change.

Example. Dealing once more with the same example as in the previous sections, we have to find a biquadratic equation

$$\tfrac{7}{8}(1-\tfrac{33}{35})\,x^4+\tfrac{49}{96}[(\tfrac{33}{35}-\tfrac{218}{245})+(1-\tfrac{33}{35})\,u_1]\,x^3+\tfrac{5}{32}[(\tfrac{218}{245}-\tfrac{62}{75})+(\tfrac{33}{35}-\tfrac{218}{245})\,u_1+(1-\tfrac{33}{35})\,u_2]\,x^2$$
$$+\tfrac{13}{384}[(\tfrac{62}{75}-\tfrac{4}{5})+(\tfrac{218}{245}-\tfrac{62}{75})\,u_1+(\tfrac{33}{35}-\tfrac{218}{245})\,u_2+(1-\tfrac{33}{35})\,u_3]\,x$$
$$+\tfrac{1}{384}[(\tfrac{4}{5}-\tfrac{4}{5})+(\tfrac{62}{75}-\tfrac{4}{5})\,u_1+(\tfrac{218}{245}-\tfrac{62}{75})\,u_2+(\tfrac{33}{35}-\tfrac{218}{245})\,u_3+(1-\tfrac{33}{35})\,u_4] = 0,$$

or, if we use four significant figures in every fraction,

$$x^4+(0\cdot5417+0\cdot5833u_1)x^3+(0\cdot1973+0\cdot1658u_1+0\cdot1786u_2)x^2$$
$$+(0\cdot01805+0\cdot04274u_1+0\cdot03593u_2+0\cdot03869u_3)x$$
$$+(0+0\cdot001389u_1+0\cdot003288u_2+0\cdot002764u_3+0\cdot002976u_4)=0.$$

This biquadratic equation must have the roots $-\frac{1}{4}\pm\sqrt{-\frac{5}{48}}$ and $-\frac{1}{8}$. If the fourth root is called $(-z)$, then the equation must be identical with

$$(x^3+\tfrac{5}{8}x^2+\tfrac{11}{48}x+\tfrac{1}{48})(x+z)=0.$$

Simple arithmetic shows then that

$$u_1=0\cdot1429+1\cdot7142z,\quad u_2=0\cdot0459+1\cdot9082z,$$
$$u_3=-0\cdot1283+2\cdot2566z \quad\text{and}\quad u_4=0\cdot0019+1\cdot9962z.$$

Now z must be at least $0\cdot05685$ to make u_3 positive and it must not exceed $0\cdot5$, because otherwise the u_i would exceed unity. But then u_4 will always be larger than u_3, unless we put $z=\frac{1}{2}$ which would mean $u_i=1$ for all i and then there would be no members at all in grades 3 and above. It follows that we must admit reversions from grade 3 into grade 2. We can then, for instance, take $z=0\cdot2$ and have

$$u_1=0\cdot4857,\quad u_2=0\cdot4275,\quad u_3=0\cdot3230\quad\text{and finally}\quad u_4=0\cdot4011.$$

The biquadratic equation becomes

$$x^4+\tfrac{33}{40}x^3+\tfrac{17}{48}x^2+\tfrac{1}{15}x+\tfrac{1}{240}=0.$$

This is the same as the one used in § 4, and we can again write down the changing pattern of the population, but this time taking also seniority into account:

Grade 2

Age	$t=0$						1					2					3				
1	73	—	—	—	—	73	43	—	—	—	43	50	—	—	—	50	52	—	—	—	52
2	12	6	—	—	—	18	40	20	—	—	60	23	12	—	—	35	27	14	—	—	41
3	3	2	1	—	—	6	4	2	2	—	8	15	6	5	—	26	8	3	3	—	14
4	3	3	3	2	—	11	—	—	—	—	0	—	1	—	—	1	1	2	1	1	5
5	—	—	2	1	1	4	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—
	91	11	6	3	1	112	87	22	3	—	112	88	19	5	—	112	88	19	4	1	112

Age	4						5					6					7 and later				
1	49	—	—	—	—	49	50	—	—	—	50	50	—	—	—	50	50	—	—	—	50
2	28	15	—	—	—	43	27	13	—	—	40	27	14	—	—	41	27	14	—	—	41
3	10	4	4	—	—	18	10	5	4	—	19	10	4	4	—	18	10	4	4	—	18
4	1	1	—	—	—	2	1	1	1	—	3	1	1	1	—	3	1	1	1	—	3
5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	88	20	4	—	—	112	88	19	5	—	112	88	19	5	—	112	88	19	5	—	112

Grade 3

Age	0						1					2					3				
2	7	—	—	—	—	7	22	—	—	—	22	13	—	—	—	13	15	—	—	—	15
3	1	2	—	—	—	3	2	2	—	—	4	6	8	—	—	14	4	5	—	—	9
4	4	3	5	—	—	12	1	—	1	—	2	—	1	1	—	2	1	2	2	—	5
5	1	2	2	2	—	7	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—
	13	7	7	2	—	29	25	2	1	1	29	19	9	1	—	29	20	7	2	—	29
	4						5					6					7 and later				
2	15	—	—	—	—	15	16	—	—	—	16	15	—	—	—	15	15	—	—	—	15
3	4	5	—	—	—	9	4	5	—	—	9	4	5	—	—	9	4	5	—	—	9
4	1	1	2	—	—	4	1	1	2	—	4	1	1	2	—	4	1	1	2	—	4
5	—	1	—	—	—	1	—	—	—	—	—	—	1	—	—	1	—	1	—	—	1
	20	7	2	—	—	29	21	6	2	—	29	20	7	2	—	29	20	7	2	—	29

We find, as before, that further splitting up of grades is impossible, if the total in each grade is to remain constant throughout the years.

Summary

This investigation deals with a stratified population, which is subject to (i) mortality, dependent on age, and to (ii) promotion rates, indicating the ratios of members of a grade which are transferred to the next higher grade at the end of the year.

Section 1 concerns a population which is not yet stratified and formulae are deduced to calculate the number of entrants at time t, necessary to replace yearly deaths and thus to keep the total of the population constant. This number depends clearly on the mortality rates and on the age distribution existing at time $t = 0$. In general the population tends towards a limiting age distribution, the 'intrinsic stationary population'.

Section 2 considers two populations and conditions are derived for the case that they need, every year, equal numbers of entrants to keep them constant.

Section 3 introduces the stratified population. Both mortality and promotion rates depend on the age, and they are independent of the time t. Under certain conditions one of the two populations considered in § 2 can be taken as the whole and the other as the lowest grade in it. It is shown how and when entries into the grade can, at the same time, replace both losses due to mortality in the whole population, and to mortality and promotion depleting the lowest grade. This can also be described by saying that the totals of both grades can be kept constant at the same time, although the age distributions change from year to year.

Section 4 generalizes the results of the previous section for a population consisting of k grades. If the population is spread over n ages, then it is shown that up to $n-1$ grades

may be possible in the most favourable case, such that they are all kept constant, whilst the age distributions all oscillate. Such a population is called semi-stationary.

Section 5 introduces the case which has been of actual importance in practical establishment work: the promotion rates are made dependent on the time spent in the grade instead of on the age.

A numerical example is attached to § 1 and is carried through all stages to illustrate the results which emerge gradually in the subsequent sections.

REFERENCES

Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, **33**, 183.

Seal, H. L. (1945). The mathematics of a population composed of k stationary strata. *Biometrika*, **33**, 226.

A SIMPLE APPROACH TO CONFOUNDING AND FRACTIONAL REPLICATION IN FACTORIAL EXPERIMENTS

By O. KEMPTHORNE, *Rothamsted Experimental Station*

Introduction

The design and analysis of factorial experiments was described in 1937 by Yates in considerable detail. In his treatment Yates described first the 2^n system and then went on to deal with 3^n experiments and experiments of the 2^m3^n type. The 2^n system is capable of very easy explanation, but with experiments of higher order both the design and analysis become of increasing complexity. It is the purpose of this paper to present a general method by which factorial designs of the type p^n may be examined, in respect of both confounding and fractional replication. The method will be described by explanation of the rules for the 2^n and 3^n systems and corresponds quite closely to that given by Fisher (1942). The present approach presents confounding and fractional replication as different aspects of the same process. Experimental designs suggested by Plackett & Burman (1946) are also discussed.

The 2^n system

In this system all combinations of n factors each at two levels are tested. The totality of treatment combinations may be represented by the points of an n-dimensional lattice, each side being of unit length. Let the factors be $x_1, x_2, \ldots, x_n$ and take n mutually orthogonal axes $y_1 \ldots y_n$. The point $(000 \ldots 0)$ will then represent the control treatment, $(1000 \ldots 0)$ the treatment consisting of x_1 at the upper level and all the other factors at the lower level, and so on. The treatment effect of x_1 is the difference of the means of the yields of plots receiving x_1 and those not receiving x_1. It is therefore the difference between the mean of the plots represented by points lying on the plane $y_1 = 1$ and the mean of those represented by the points on the plane $y_1 = 0$. The interaction of x_1 and x_2 is the difference between the means of those plots represented by $y_1 = 1$, $y_2 = 1$ or $y_1 = 0$, $y_2 = 0$ and those represented by $y_1 = 0$, $y_2 = 1$ and $y_1 = 1$, $y_2 = 0$, i.e. the difference of the means of those plots for which

$$y_1 + y_2 = 2 \text{ or } = 0 \pmod 2,$$

and those for which $$y_1 + y_2 = 1 \pmod 2.$$

Similarly, the triple interaction of x_1, x_2 and x_3 is the difference between the means of those plots for which $$y_1 + y_2 + y_3 = 0 \pmod 2,$$

and those for which $$y_1 + y_2 + y_3 = 1 \pmod 2.$$

This process can be continued to the consideration of the interaction of $x_1, x_2, \ldots, x_n$ which is the difference between the mean of those plots for which

$$y_1 + y_2 + y_3 + \ldots + y_n = 0 \pmod 2,$$

and the mean of those for which

$$y_1 + y_2 + y_3 + \ldots + y_n = 1 \pmod 2.$$

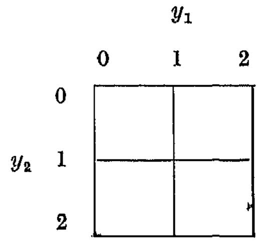
In the n-dimensional space parallel hyper-planes may be drawn containing the points of the lattice, such that the total yield forming the positive part of an interaction is obtained from

a set of parallel hyper-planes equidistant from each other. Likewise the negative part is obtained from another set of parallel hyper-planes, each plane of which lies midway between two planes of the first set.

The 3^n system

With n factors at each of three levels the treatment combinations are given by an n-dimensional lattice, each side being of length two units and containing three points. The treatment contrasts may be described as in the 2^n system with some slight modifications.

Any contrast in the 3^n system involves the comparison of three totals of the yields of 3^{n-1} plots, and may be represented by the comparison of the differences between the yields of the plots lying on three sets of parallel hyper-planes. For example, if $n = 2$ the lattice is as follows:



The main effect of x_1 is the difference between the totals of yields of plots for which $y_1 = 0$, $y_1 = 1$ and $y_1 = 2$. The I component* of the interaction of x_1 and x_2 is the difference between the totals of the yields of plots for which

$$y_1 - y_2 = 0, \quad y_1 - y_2 = 1, \quad \text{and} \quad y_1 - y_2 = 2.$$

The J component is given by the contrast between the yields of plots for which

$$y_1 + y_2 = 0, \quad y_1 + y_2 = 1, \quad \text{and} \quad y_1 + y_2 = 2.$$

Anticipating the extension to cases when n is greater than 2, the equations for the I component may be written as follows:

$$X_1X_2(I_0): \ y_1 + 2y_2 = 0 \pmod 3,$$
$$X_1X_2(I_1): \ y_1 + 2y_2 = 1 \pmod 3,$$
$$X_1X_2(I_2): \ y_1 + 2y_2 = 2 \pmod 3.$$

If x_1 and x_2 (and therefore y_1 and y_2) are interchanged, then $X_2X_1(I_0)$ is given by the equations $y_2 + 2y_1 = 0$, $X_2X_1(I_1)$ by $y_2 + 2y_1 = 1$, $X_2X_1(I_2)$ by $y_2 + 2y_1 = 2$, all mod 3. But the equation $y_2 + 2y_1 = 0 \pmod 3$ is identical with the equation $y_1 + 2y_2 = 0 \pmod 3$, since $3y_1 + 3y_2 = 0 \pmod 3$, whatever the values of y_1 and y_2; $X_2X_1(I_0)$ is therefore equal to $X_1X_2(I_0)$. Subtracting the equation $y_2 + 2y_1 = 1 \pmod 3$ from the equation

$$3y_1 + 3y_2 = 0 = 3 \pmod 3,$$

we get $y_1 + 2y_2 = 2 \pmod 3$; $X_2X_1(I_1)$ is therefore identical with $X_1X_2(I_2)$. It is obvious from the equations given above for the J component that $X_1X_2(J_i) = X_2X_1(J_i)$ for $i = 0$, 1 and 2.

* Yates's terminology for the components of interactions is used where convenient, but it is more convenient to refer to I_1, I_2 and I_3 as I_0, I_1 and I_2 respectively.

Considering the case $n = 3$, it is easily seen that the second order interaction may be split into four parts each consisting of the contrasts between three totals. These may be represented by the following equations:

$$\begin{array}{ll}
\text{(I)} & y_1+\ y_2+\ y_3 = 0 \pmod 3, \\
& y_1+\ y_2+\ y_3 = 1 \\
& y_1+\ y_2+\ y_3 = 2 \\
\text{(II)} & y_1+2y_2+\ y_3 = 0 \pmod 3, \\
& y_1+2y_2+\ y_3 = 1 \\
& y_1+2y_2+\ y_3 = 2 \\
\text{(III)} & y_1+\ y_2+2y_3 = 0 \pmod 3, \\
& y_1+\ y_2+2y_3 = 1 \\
& y_1+\ y_2+2y_3 = 2 \\
\text{(IV)} & y_1+2y_2+2y_3 = 0 \pmod 3. \\
& y_1+2y_2+2y_3 = 1 \\
& y_1+2y_2+2y_3 = 2
\end{array}$$

In order these have been named by Yates

$$Z,\quad X,\quad Y,\quad W.$$

It is interesting in passing to note the relations between Z, X, Y and W for permutations of the order of the factors. It is obvious from (I) that Z is invariant for any change in order of the three factors X_1, X_2 and X_3. Interchanging y_2 and y_3, equations (II) become equations (III), so that $ABC(X) = ACB(Y)$. The following interchanges may be easily verified (using the equation $3y_1+3y_2+3y_3 = 0 \pmod 3$ where necessary):

$$ABC(X) = BCA(Y) = CAB(Y) = ACB(Y) = CBA(X) = BAC(W).$$

From the equations, it is clear that Z, X and W may be computed in the way given by Yates, since

$$Z = J\{x_1, J(x_2,x_3)\}, \qquad Y = J\{x_1, I(x_2,x_3)\},$$
$$X = I\{x_1, I(x_2,x_3)\}, \qquad W = I\{x_1, J(x_2,x_3)\},$$

$I(x_2,x_3)$ and $J(x_2,x_3)$ being evaluated for each level of x_1. The extension to the case $n = 4$ is again obvious; the main effects, two-factor and three-factor interactions, follow as in the above, and the four-factor interaction may be split into eight comparisons of three totals:

$$\begin{array}{rll}
\text{I} & y_1+\ y_2+\ y_3+\ y_4 = 0,1,2 & \pmod 3, \\
\text{II} & y_1+\ y_2+\ y_3+2y_4 = 0,1,2 & \pmod 3, \\
\text{III} & y_1+\ y_2+2y_3+\ y_4 = 0,1,2 & \pmod 3, \\
\text{IV} & y_1+\ y_2+2y_3+2y_4 = 0,1,2 & \pmod 3, \\
\text{V} & y_1+2y_2+\ y_3+\ y_4 = 0,1,2 & \pmod 3, \\
\text{VI} & y_1+2y_2+\ y_3+2y_4 = 0,1,2 & \pmod 3, \\
\text{VII} & y_1+2y_2+2y_3+\ y_4 = 0,1,2 & \pmod 3, \\
\text{VIII} & y_1+2y_2+2y_3+2y_4 = 0,1,2 & \pmod 3.
\end{array}$$

As in the case of two factors, the effect of permutations of the order on the components of X, Y, Z and W may be easily obtained. The four-factor interactions may be computed by putting the equations given above into the following form:

$$\begin{aligned} \mathrm{I} &= J\{x_1, Z\}, & \mathrm{V} &= I\{x_1, W\},\\ \mathrm{II} &= J\{x_1, Y\}, & \mathrm{VI} &= I\{x_1, X\},\\ \mathrm{III} &= J\{x_1, X\}, & \mathrm{VII} &= I\{x_1, Y\},\\ \mathrm{IV} &= J\{x_1, W\}, & \mathrm{VIII} &= I\{x_1, Z\}, \end{aligned}$$

where the three components of W, X, Y, Z of x_2, x_3, x_4 (in that order) are evaluated for each level of x_1.

The p^n system

The total of p^n-1 degrees of freedom, where p is a prime, in the analysis of variance of a p^n experiment may be split into $(p^n-1)/(p-1)$ sets of $(p-1)$ degrees of freedom, the contrasts being given by the following hyper-planes:

$$\text{Main effects}\quad \begin{array}{l} y_1 = 0, 1, 2, \ldots, p-1,\\ y_2 = 0, 1, 2, \ldots, p-1.\\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ y_p = 0, 1, 2, \ldots, p-1. \end{array} \quad (\mathrm{mod}\ p),$$

$$\begin{array}{r} \text{Interactions of pairs of}\\ \text{factors, e.g. of } x_1 \text{ and } x_2 \end{array} \left\{ \begin{array}{r} y_1+y_2 = 0, 1, 2, \ldots, p-1,\\ y_1+2y_2 = 0, 1, 2, \ldots, p-1,\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_1+(p-1)\,y_2 = 0, 1, 2, \ldots, p-1. \end{array} \right. \quad (\mathrm{mod}\ p),$$

and so on to the interaction between all the factors which is given by the hyper-planes

$$a_1y_1+a_2y_2+a_3y_3+\ldots+a_ny_n = 0, 1, 2, \ldots, p-1 \quad (\mathrm{mod}\ p),$$

where a_1 equals 1 and $a_2, a_3, \ldots, a_n$ each may take all values from 1 to $p-1$.

Simplification of notation

The p^n-1 degrees of freedom in the p^n system may be split into $(p^n-1)/(p-1)$ sets of $(p-1)$ degrees of freedom, given by the above hyper-planes, but it is only necessary to specify one hyper-plane of each set of the parallel hyper-planes.

All the comparisons may be denoted by $y_1^{\alpha_1}y_2^{\alpha_2}, \ldots, y_n^{\alpha_n}$, the symbol meaning that the comparisons are given by the hyper-planes

$$\alpha_1y_1+\alpha_2y_2+\ldots+\alpha_ny_n = 0, 1, 2, \ldots, p-1 \quad (\mathrm{mod}\ p).$$

In order to obtain an enumeration which covers all the possibilities once and once only, it is necessary to use the rule that the factors are always written down in ascending order—i.e. $y_i^{\alpha_i}y_j^{\alpha_j}y_k^{\alpha_k}$, etc., such that $i<j<k\ldots$ and that $\alpha_i = 1$.

THE 3^n SYSTEM IN THE REVISED NOTATION

As an example, the 3^3 system will be examined in detail. The effects are represented by y_1, y_2, y_3; interactions between pairs, $y_1y_2, y_1y_2^2, y_1y_3, y_1y_3^2, y_2y_3, y_2y_3^2$; interactions between all three factors, $y_1y_2y_3$, $y_1y_2y_3^2$, $y_1y_2^2y_3$, $y_1y_2^2y_3^2$. Any other combination of powers of the y's can be reduced to the above set.

It is interesting to examine the interactions of the effects and interactions. In the case of the 2^n system, Yates refers to the generalized interaction of two interactions $ABCD$ and CDE say, which is ABE. The interaction of effects or interactions A and B consists of AB and AB^2 in the 3^n system.

(*a*) The interactions of main effects are obviously interactions between pairs of factors.

(*b*) The interactions of main effects and two-factor interactions with one letter in common are *two-factor interactions and main effects: e.g. the interactions of* y_1 and y_1y_2 are

$$y_1^2y_2 = y_1y_2^2, \quad \text{and} \quad y_1^3y_2^2 = y_2,$$

and the interactions of y_1 and $y_1y_2^2$ are $y_1^2y_2^2 = y_1y_2$, and $y_1^3y_2^4 = y_2$.

(*c*) The interaction between main effects and three-factor interaction are two-factor and three-factor interactions:

Between	Interactions
y_1 and $y_1y_2y_3$	$y_1^2y_2y_3 = y_1y_2^2y_3^2$, $y_1^3y_2^2y_3^2 = y_2y_3$
y_1 and $y_1y_2y_3^2$	$y_1^2y_2y_3^2 = y_1y_2^2y_3$, $y_1^3y_2^2y_3^4 = y_2y_3^2$
y_1 and $y_1y_2^2y_3$	$y_1^2y_2^2y_3 = y_1y_2y_3^2$, $y_1^3y_2^4y_3^2 = y_2y_3^2$
y_1 and $y_1y_2^2y_3^2$	$y_1^2y_2^2y_3^2 = y_1y_2y_3$, $y_1^3y_2^4y_3^4 = y_2y_3$

(*d*) The interaction between two-factor interactions are exemplified in the following table:

Between	Interactions	
y_1y_2 and $y_1y_2^2$	$y_1^2y_2^3 = y_1$,	$y_1^3y_2^5 \quad = y_2$
y_1y_2 and y_2y_3	$y_1y_2^2y_3$,	$y_1y_2^3y_3^2 = y_1y_3^2$
y_1y_2 and $y_2y_3^2$	$y_1y_2^2y_3^2$,	$y_1y_2^3y_3^4 = y_1y_3$

(*e*) The interaction between two-factor and three-factor interactions are exemplified in the following table:

Between / and	y_1y_2		$y_1y_2^2$	
$y_1y_2y_3$	$y_1y_2y_3^2$	y_3	$y_1y_3^2$	$y_2y_3^2$
$y_1y_2y_3^2$	$y_1y_2y_3$	y_3	y_1y_3	y_2y_3
$y_1y_2^2y_3$	$y_1y_3^2$	y_2y_3	$y_1y_2^2y_3^2$	y_3
$y_1y_2^2y_3^2$	y_1y_3	$y_2y_3^2$	$y_1y_2^2y_3$	y_3

The interactions between two-factor and three-factor interactions are therefore two-factor interactions in some cases and main effects and three-factor interactions in the other cases.

(*f*) The interactions between three-factor interactions are set out in the following diagram:

Between / and	$y_1y_2y_3$	$y_1y_2y_3^2$	$y_1y_2^2y_3$	$y_1y_2^2y_3^2$
$y_1y_2y_3$	— —	y_1y_2, y_3	y_1y_3, y_2	y_1, y_2y_3
$y_1y_2y_3^2$		— —	y_1, $y_2y_3^2$	$y_1y_3^2$, y_2
$y_1y_2^2y_3$			— —	$y_1y_2^2$, y_3
$y_1y_2^2y_3^2$				— —

Confounding

Confounding or the allocation of treatment combinations to blocks implies the allocation of all the points of the lattice into p^c sets, of p^{n-c} points, such that the comparisons between these sets involve particular sets of $p-1$ degrees of freedom. The aim of confounding is to reduce the effect of soil heterogeneity by reducing block size, but ensuring that the block comparisons have little possible practical importance.

If comparisons $A = y_1^{\alpha_1}y_2^{\alpha_2}\dots y_n^{\alpha_n}$ and $B = y_1^{\beta_1}y_2^{\beta_2}\dots y_n^{\beta_n}$ are confounded, then so is their generalized interaction, i.e. all the products of these two, i.e. AB, $AB^2, \dots, AB^{p-1}$. For, if the treatment combinations for which $\alpha_1 y_1+\alpha_2 y_2+\dots+\alpha_n y_n$ is equal to $0, 1, 2, \dots, p-1$ are put into separate blocks and also those treatments for which $\beta_1 y_1+\beta_2 y_2+\dots+\beta_n y_n$ is equal to $0, 1, 2, \dots, p-1$, then $(\alpha_1+\lambda\beta_1)y_1+(\alpha_2+\lambda\beta_2)y_2+\dots+(\alpha_n+\lambda\beta_n)y_n$ is equal (mod p) to $0, 1, 2, \dots, p-1$ for all λ from 0 to $p-1$.

The present approach to confounding of the 2^n system is identical with that given by Yates and we proceed to consider the rather more complex case of the 3^n system.

(*a*) 3^3 *system*

(1) *In blocks of* 3^2. Any three-factor interaction may be confounded.

(2) *In blocks of* 3. We cannot confine the confounded degrees of freedom to three-factor interactions because the generalized interaction of any two reduces to a two-factor interaction and a main effect. If two three-factor interactions, $y_1y_2y_3$ and $y_1y_2^2y_3$ are confounded, the 8 degrees of freedom for blocks may be described as follows:

	D.F.
y_2	2
y_1y_3	2
$y_1y_2y_3$	2
$y_1y_2^2y_3$	2
	8

We can, however, choose three two-factor interactions and one three-factor interaction pair for our block comparisons.

(b) 3^4 system in blocks of 3^2

It is immediately obvious that we can confound two two-factor interactions and two higher-order interaction pairs to give blocks of nine. The important point, however, is to find a design confounding only three-factor interaction pairs.

We therefore evaluate the interactions of all pairs of three-factor interactions, which have two letters in common. These may be derived from the interaction of $y_1y_2y_3$ with the four three-factor interactions of y_1, y_2 and y_4, which are as follows:

Interaction of $y_1y_2y_3$	and	$y_1y_2y_4$	$y_1y_2y_3^2y_4^2$	and	$y_3y_4^2$,
Interaction of $y_1y_2y_3$	and	$y_1y_2y_4^2$	$y_1y_2y_3^2y_4$	and	y_3y_4,
Interaction of $y_1y_2y_3$	and	$y_1y_2^2y_4$	$y_1y_3^2y_4^2$	and	$y_2y_3^2y_4$,
Interaction of $y_1y_2y_3$	and	$y_1y_2^2y_4^2$	$y_1y_3^2y_4$	and	$y_2y_3^2y_4^2$.

Obviously there are many designs for the 3^4 design in nine blocks of nine plots confounding three-factor interactions. Those which confound four-factor interactions must also confound two-factor interactions. The names of the confounded interactions and their squares (each of which corresponds to the same grouping as the element itself) form a group with the identity and the equation $y_i^3 = 1$, for all i, and further work is presumably most promising on these lines.

(c) 3^5 in blocks of 9

There is no design confounding only three-factor or higher-order interactions. If one two-factor interaction can be sacrificed, a possible scheme of confounding is given by the following table of generalized interactions:

Between / and	$y_1y_2y_3$	$y_1y_2^2y_4^2$	$y_1y_3^2y_4$	$y_2y_3^2y_4^2$
$y_3y_4y_5$	$y_1y_2y_3^2y_4y_5$ $y_1y_2y_4^2y_5^2$	$y_1y_2^2y_3y_5$ $y_1y_2^2y_3^2y_4y_5^2$	$y_1y_4^2y_5$ $y_1y_3y_5^2$	y_2y_5 $y_2y_3y_4y_5^2$

This two-factor interaction is estimated by the comparison of three sets of nine blocks, and the accuracy of the estimate will be low.

(d) 3^6 in blocks of 27

We may, for example, confound the following:

	$y_2y_4y_6$	
$y_1y_2y_3$	$y_1y_2^2y_3y_4y_6$	$y_1y_3y_4^2y_6^2$
$y_1y_4y_5$	$y_1y_2y_4^2y_5y_6$	$y_1y_2^2y_5y_6^2$
$y_1y_2^2y_3^2y_4^2y_5^2$	$y_1y_3^2y_5^2y_6$	$y_1y_2y_3^2y_4y_5^2y_6^2$
$y_2y_3y_4^2y_5^2$	$y_2y_3^2y_5y_6^2$	$y_3y_4y_5^2y_6^2$

Three three-factor interactions, six four-factor interactions, three five-factor interactions and one six-factor interaction are confounded. If y_6 is omitted from all the above expressions

we obtain a 3^5 experiment in blocks of nine confounding one two-factor interaction, seven three-factor interactions, three four-factor and two five-factor interactions—that is, the design given above for the 3^5 system.

Extension to more complicated cases

Extensions of the above to more complicated cases should most easily be achieved by the use of group theory. The confounding of a p^n design in p^c blocks corresponds to a group of $\frac{1}{2}(p^c+1)$ elements such that all except the unit element involve at least a certain number of letters. For most agricultural experiments each element should contain at least three letters, so that no main effects or two-factor interactions are confounded. The group is an Abelian group and if A and B are elements of the group so are $AB, AB^2, \ldots, AB^{p-1}$. The order of each element is p, and if A is an element so are the first $(p-1)$ powers of A. This aspect is being followed, and it is hoped will yield results.

Fractional replication in the 2^n system

Some principles of fractional replication have been worked out over the past few years at Rothamsted (Finney, 1945). In the case of a 2^n system, with factors $a_1 \ldots a_n$ say, a half-replicate might consist of those treatment combinations which form the positive part of the interaction $A_1A_2 \ldots A_n$. Each function of the plot yields consisting of the sum of one-half of them minus the sum of the other half then corresponds to two degrees of freedom. Alternatively, each degree of freedom has one alias, and the aim in fractional replication is to design the experiment so that the aliases of effects which the experimenter wishes to measure are high-order interactions which could not possibly have practical significance.

For convenience of presentation, we develop first the theory for the case of the 2^n system. Suppose that of all the points on the lattice for the 2^n system, only those points for which

$$y_1+y_2+y_3+\ldots+y_n = 0$$

are included in the experiment. Then the points on the hyper-plane $y_1 = 0$, also lie on the plane $y_2+y_3+\ldots+y_n = 0$; and likewise those for which $y_1 = 1$ lie on the plane

$$y_2+y_3+\ldots+y_n = 1.$$

The contrast which we have denoted by y_1 is therefore identical with that denoted by $y_2y_3y_4\ldots y_n$. Again, if we suppose that only those treatment combinations are tested which lie on the hyper-planes

$$\alpha_1y_1+\alpha_2y_2+\ldots+\alpha_ny_n = 0, \quad \beta_1y_1+\beta_2y_2+\ldots+\beta_ny_n = 0,$$

then the points will also lie on the intersection of these planes which is given by the equation

$$(\alpha_1+\beta_1)y_1+(\alpha_2+\beta_2)y_2+\ldots+(\alpha_n+\beta_n) = 0 \pmod 2.$$

The points which lie on the hyper-planes

$$\gamma_1y_1+\gamma_2y_2+\ldots+\gamma_ny_n = 0, 1 \pmod 2$$

will also lie on the planes

$$\begin{aligned}
(\alpha_1+\gamma_1)y_1+(\alpha_2+\gamma_2)y_2 \quad &+\ldots+ \quad (\alpha_n+\gamma_n)y_n = 0, 1 \pmod 2,\\
(\beta_1+\gamma_1)y_1+(\beta_2+\gamma_2)y_2 \quad &+\ldots+ \quad (\beta_n+\gamma_n)y_n = 0, 1 \pmod 2,\\
(\alpha_1+\beta_1+\gamma_1)y_1+(\alpha_2+\beta_2+\gamma_2)y_2 &+\ldots+(\alpha_n+\beta_n+\gamma_n)y_n = 0, 1 \pmod 2.
\end{aligned}$$

Changing to the simpler notation, these results may be obtained by equating to unity the symbols corresponding to the effects which the experiment cannot measure (as only treat-

ment combinations of the same sign in the function giving the effect are included) and multiplying the symbol corresponding to a particular effect by these symbols. Thus we put

$$I = y_1^{\alpha_1} y_2^{\alpha_2} \dots y_n^{\alpha_n} = y_1^{\beta_1} y_2^{\beta_2} \dots y_n^{\beta_n} = y_1^{\alpha_1+\beta_1} y_2^{\alpha_2+\beta_2} \dots y_n^{\alpha_n+\beta_n},$$

then the contrast

$$y_1^{\gamma_1} y_2^{\gamma_2} \dots y_n^{\gamma_n}$$

is the same as those given by

$$y_1^{\alpha_1+\gamma_1} y_2^{\alpha_2+\gamma_2} \dots y_n^{\alpha_n+\gamma_n}, \quad y_1^{\beta_1+\gamma_1} y_2^{\beta_2+\gamma_2} \dots y_n^{\beta_n+\gamma_n} \quad \text{and} \quad y_1^{\alpha_1+\beta_1+\gamma_1} y_2^{\alpha_2+\beta_2+\gamma_2} \dots y_n^{\alpha_n+\beta_n+\gamma_n},$$

where each power is reduced modulus 2.

2^n SYSTEM WITHOUT SUBDIVISION INTO BLOCKS

We now consider some of the possibilities of partial replication for the 2^n system. The basis of designs with fractional replication is the choice of an identity relationship; most of the possible relationships are of no value, and we consider only those which yield the least possible confusion between main effects and first-order interactions.

Half-replication

$n = 3$. If we take $I = y_1 y_2 y_3$, then $y_1 = y_1(y_1 y_2 y_3) = y_1^2 y_2 y_3 = y_2 y_3$. Such a design which confuses main effects and two-factor interactions would not be of any practical use.

$n = 4$. If we take $I = y_1 y_2 y_3 y_4$, then the aliases are exemplified by

$$y_1 = y_2 y_3 y_4 \quad \text{and} \quad y_1 y_2 = y_3 y_4.$$

Such a design would not be used unless the experimenter were confident that two-factor interactions were negligible.

$n = 5$. If we take $I = y_1 y_2 y_3 y_4 y_5$, then the aliases are exemplified by

$$y_1 = y_2 y_3 y_4 y_5 \quad \text{and} \quad y_1 y_2 = y_3 y_4 y_5.$$

A half-replicate with five or more factors is feasible when there is no necessity to remove heterogeneity by the use of blocks, since main effects will have aliases which are interactions of four factors at least, and two-factor interactions will have aliases which are interactions of at least three factors.

Quarter-replication

Each degree of freedom will now have three aliases. For each value of n we give the identity relationship and typical alias relationships.

$n = 4$. $\quad I = y_1 y_2 = y_3 y_4 = y_1 y_2 y_3 y_4$;

then $\quad y_1 = y_2 = y_1 y_3 y_4 = y_2 y_3 y_4 \quad$ and $\quad y_1 y_3 = y_2 y_3 = y_1 y_4 = y_2 y_4$.

$n = 5$. $\quad I = y_1 y_2 = y_3 y_4 y_5 = y_1 y_2 y_3 y_4 y_5 \quad (a)$,

or $\quad I = y_1 y_2 y_3 = y_3 y_4 y_5 = y_1 y_2 y_4 y_5 \quad (b)$.

(*a*) Gives $\quad y_1 = y_2 = y_1 y_3 y_4 y_5 = y_2 y_3 y_4 y_5 \quad$ and $\quad y_1 y_3 = y_2 y_3 = y_1 y_4 y_5 = y_2 y_4 y_5$.

(*b*) Gives $\quad y_1 = y_2 y_3 = y_1 y_3 y_4 y_5 = y_2 y_4 y_5$.

$n = 6$. $\quad I = y_1 y_2 y_3 y_4 = y_3 y_4 y_5 y_6 = y_1 y_2 y_5 y_6$;

then $\quad y_1 = y_2 y_3 y_4 = y_1 y_3 y_4 y_5 y_6 = y_2 y_5 y_6$

and $\quad y_1 y_2 = y_3 y_4 = y_1 y_2 y_3 y_4 y_5 y_6 = y_5 y_6$.

$n = 7$. $\quad I = y_1 y_2 y_3 y_4 = y_4 y_5 y_6 y_7 = y_1 y_2 y_3 y_5 y_6 y_7$;

then $\quad y_1 = y_2 y_3 y_4 \quad$ and $\quad y_1 y_2 = y_3 y_4$.

$n = 8$. $\quad I = y_1 y_2 y_3 y_4 y_5 = y_4 y_5 y_6 y_7 y_8 = y_1 y_2 y_3 y_6 y_7 y_8$;

then $\quad y_1 = y_2 y_3 y_4 y_5 \quad$ and $\quad y_1 y_2 = y_3 y_4 y_5$.

Designs in quarter replicate are therefore possible when n is greater than or equal to 8.

High-order fractional replication

In general, the existence of fractional designs of the 2^n system with fraction 2^p, which will be useful where information on all main effects and two-factor interactions is required, depends on the existence of a group of 2^p elements, one element being unity and the other elements all containing at least five letters. No simple method has been found of enumerating such groups, but it is perhaps worth recording the following designs which appear to represent the greatest degree of fractional replication possible.

(*a*) *Eighth replication*

If we are testing ten or more factors at each of two levels, one-eighth of a replication will enable main effects and two-factor interactions to be estimated. An appropriate identity relationship is the following:

$$\begin{aligned} I &= y_1y_2y_3y_4y_5 = y_1y_2y_6y_7y_8 = y_3y_4y_5y_6y_7y_8 \\ &= y_1y_3y_7y_9y_{10} = y_2y_4y_5y_7y_9y_{10} = y_2y_3y_6y_8y_9y_{10} = y_1y_4y_5y_6y_8y_9y_{10}. \end{aligned}$$

Thus ten main effects and forty-five two-factor interactions may be estimated from a trial testing 128 of the 1024 possible treatment combinations.

(*b*) *Sixteenth replication*

If we are testing twelve or more factors a possible identity relationship is the following:

$$\begin{aligned} I &= y_1y_2y_3y_4y_5 = y_1y_2y_6y_7y_8 = y_3y_4y_5y_6y_7y_8 = y_1y_2y_9y_{10}y_{11} \\ &= y_3y_4y_5y_9y_{10}y_{11} = y_6y_7y_8y_9y_{10}y_{11} = y_1y_2y_3y_4y_5y_6y_7y_8y_9y_{10}y_{11} \\ &= y_1y_3y_6y_9y_{12} = y_2y_4y_5y_6y_9y_{12} = y_2y_3y_7y_8y_9y_{12} = y_1y_4y_5y_7y_8y_9y_{12} \\ &= y_2y_3y_6y_{10}y_{11}y_{12} = y_1y_4y_5y_6y_{10}y_{11}y_{12} = y_1y_3y_7y_8y_{10}y_{11}y_{12} = y_2y_4y_5y_7y_8y_{10}y_{11}y_{12}. \end{aligned}$$

In this case twelve main effects and sixty-six two-factor interactions may be estimated from a trial testing 256 of the possible 4096 treatment combinations.

The extent to which these designs will be of practical value depends very much on the existence of a sufficient mass of reasonably homogeneous material to test the large number of treatment combinations without the necessity of dividing the material into smaller batches and using the device of confounding. An experiment involving say 256 different treatment combinations is not large by modern standards. At Rothamsted, for example, an experiment involving 200 distinct treatments on 300 plots has been carried out for some years: this experiment was, however, made possible by utilizing the elimination of the effects of soil heterogeneity by highly complex confounding; the design, in fact, consisted of three 5×5 lattice squares necessitating seventy-five plots, and each of these plots was split into four subplots. The advantages of testing twelve factors, say, at the same time under virtually the same experimental conditions cannot, however, be ignored. Such an experiment should have more value, other things being equal, than two distinct experiments each testing some of the factors. An examination has not been made of the possibilities of reducing block size by confounding for the above two designs, but it is probably necessary to sacrifice a few two-factor interactions.

The relationship between fractional replication and confounding

It is clear that fractional replication and confounding are different aspects of the same process. A 2^n design of 2^p blocks may be described as a 1 in 2^p replicate of a 2^{n+p} design with no subdivision into blocks, by regarding the blocks as a 2^p system in p factors. As an

example, consider the 2^5 design in y_1, y_2, y_3, y_4 and y_5 laid out in four blocks of eight and confounding $y_1y_2y_3$, $y_3y_4y_5$ and $y_1y_2y_4y_5$; superimposing two pseudo-factors b_1 and b_2, the experiment is a quarter-replicate of a 2^7 design in $y_1, y_2, y_3, y_4, y_5, b_1, b_2$. The identity on which the quarter replicate is based is given by the equations

$$b_1 = y_1y_2y_3, \quad b_2 = y_3y_4y_5, \quad b_1b_2 = y_1y_2y_4y_5$$

or the equation

$$I = y_1y_2y_3b_1 = y_3y_4y_5b_2 = y_1y_2y_4y_5b_1b_2.$$

If we examine this equation in the same way as in the previous sections, we find that the design depends on the fact that the aliases of the following type may be ignored:

$$y_1 = y_2y_3b_1 = y_1y_3y_4y_5b_2 = y_2y_4y_5b_1b_2,$$
$$y_1y_2 = y_3b_1 = y_1y_2y_3y_4y_5b_2 = y_4y_5b_1b_2.$$

This example is worth pursuing. The design is frequently used with one replication only, the error being estimated from three-factor and higher-order interactions. We set out below the identity and 31 degrees of freedom together with all their aliases and their usual place in the analysis of variance—blocks (B), treatment (T), or error (E). For convenience of printing we denote the factors tested in the experiment by a, b, c, d, e instead of $y_1y_2y_3y_4y_5$ and the block factors by x and y. Capitals are used for treatment effects thus conforming to present usage.

I	$= ABCX$	$= CDEY$	$= ABDEXY$	
A	$= BCX$	$= ACDEY$	$= BDEXY$	T
B	$= ACX$	$= BCDEY$	$= ADEXY$	T
AB	$= CX$	$= ABCDEY$	$= DEXY$	T
C	$= ABX$	$= DEY$	$= ABCDEXY$	T
AC	$= BX$	$= ADEY$	$= BCDEXY$	T
BC	$= AX$	$= BDEY$	$= ACDEXY$	T
ABC	$= X$	$= ABDEY$	$= CDEXY$	B
D	$= ABCDX$	$= CEY$	$= ABEXY$	T
AD	$= BCDX$	$= ACEY$	$= BEXY$	T
BD	$= ACDX$	$= BCEY$	$= AEXY$	T
ABD	$= CDX$	$= ABCEY$	$= EXY$	E
CD	$= ABDX$	$= EY$	$= ABCEXY$	T
ACD	$= BDX$	$= AEY$	$= BCEXY$	E
BCD	$= ADX$	$= BEY$	$= ACEXY$	E
$ABCD$	$= DX$	$= ABEY$	$= CEXY$	E
E	$= ABCEX$	$= CDY$	$= ABDXY$	T
AE	$= BCEX$	$= ACDY$	$= BDXY$	T
BE	$= ACEX$	$= BCDY$	$= ADXY$	T
ABE	$= CEX$	$= ABCDY$	$= DXY$	E
CE	$= ABEX$	$= DY$	$= ABCDXY$	T
ACE	$= BEX$	$= ADY$	$= BCDXY$	E
BCE	$= AEX$	$= BDY$	$= ACDXY$	E
$ABCE$	$= EX$	$= ABDY$	$= CDXY$	E
DE	$= ABCDEX$	$= CY$	$= ABXY$	T
ADE	$= BCDEX$	$= ACY$	$= BXY$	E
BDE	$= ACDEX$	$= BCY$	$= AXY$	E
$ABDE$	$= CDEX$	$= ABCY$	$= XY$	B
CDE	$= ABDEX$	$= Y$	$= ABCXY$	B
$ACDE$	$= BDEX$	$= AY$	$= BCXY$	E
$BCDE$	$= ADEX$	$= BY$	$= ACXY$	E
$ABCDE$	$= DEX$	$= ABY$	$= CXY$	E

If we take for each linear function of the yields the alias involving the smallest possible number of letters, but remembering that x, y are pseudo-factors, so that X, Y and XY are of

equal importance and therefore XY should be regarded as a main effect and not an interaction, we have the following allocation of contrasts to the three components of the analysis of variance:

Blocks:	$X,\ Y,\ XY.$
Treatments:	$A,\ B,\ C,\ D,\ E.$
	$AB = CX,\quad AC = BX,\quad BC = AX,$
	$CD = EY,\quad DE = CY,\quad CE = DY.$
	$AD,\ BD,\ AE,\ BE.$
Error:	$AY,\ BY,\ DX,\ EX,\ AXY,\ BXY,\ CXY,\ DXY,\ EXY,\ ACD,\ BCD,\ ACE,\ BCE.$

The four three-factor interactions could equally well be regarded as interactions between two-factor interactions and blocks. It would be anticipated that these would be smaller than the interactions of main effects and blocks. The purpose of the present exposition is to give a clear statement of the possible interpretations of the results of an individual experiment. Further remarks on the problem of interpretation are postponed to a later section in the paper.

An example of fractional replication with confounding

A design which has proved of practical utility is the half-replicate of a 2^6 experiment arranged in four blocks of eight plots.

Call the factors $y_1, y_2, y_3, y_4, y_5, y_6$. Then the best confounding is that in which, using full replication, the block differences are all third-order interactions, say

$$y_1y_2y_3y_4,\quad y_3y_4y_5y_6\quad \text{and}\quad y_1y_2y_5y_6.$$

But it is impossible to keep main effects and interactions clear with this confounding, whatever interaction is equated to the identity.

If we take the confounded interactions to be of the type

$$y_1y_2y_3,\quad y_3y_4y_5,\quad y_1y_2y_4y_5,$$

and the interaction $y_1y_2y_3y_4y_5y_6$ to be unity, then the following interactions are also confounded:

$$y_4y_5y_6,\quad y_1y_2y_6\quad \text{and}\quad y_3y_6.$$

It will be found by enumeration of the possibilities that one first-order interaction must be sacrificed. All main effects and the other first-order interactions will have high-order aliases.

It is interesting to examine this design in the same way as the 2^5 above for the relations between block-treatment interactions and treatment interactions.

There are, in fact, only thirty-two independent contrasts, and it is simplest to enumerate these by operating on the identity relationship with the thirty-two possibilities for the 2^5 system omitting y_6. As before, we insert block pseudo-factors. For simplicity of printing we use A, B, C, D, E, F for the factors and X, Y for the block factors. Then

$$I = ABCDEF,\quad X = ABC,\quad Y = CDE,\quad XY = ABDE,$$

and combining these into one relationship, we have

$$I = ABCDEF = ABCX = CDEY = ABDEXY = DEFX = ABFY = CFXY.$$

A complete table of the aliases for this design follows:

I	$= ABCDEF$	$= ABCX$	$= DEFX$	$= CDEY$	$= ABFY$	$= ABDEXY$	$= CFXY$	
A	$= BCDEF$	$= BCX$	$= ADEFX$	$= ACDEY$	$= BFY$	$= BDEXY$	$= ACFXY$	T
B	$= ACDEF$	$= ACX$	$= BDEFX$	$= BCDEY$	$= AFY$	$= ADEXY$	$= BCFXY$	T
AB	$= CDEF$	$= CX$	$= ABDEFX$	$= ABCDEY$	$= FY$	$= DEXY$	$= ABCFXY$	T
C	$= ABDEF$	$= ABX$	$= CDEFX$	$= DEY$	$= ABCFY$	$= ABCDEXY$	$= FXY$	T
AC	$= BDEF$	$= BX$	$= ACDEFX$	$= ADEY$	$= BCFY$	$= BCDEXY$	$= AFXY$	T
BC	$= ADEF$	$= AX$	$= BCDEFX$	$= BDEY$	$= ACFY$	$= ACDEXY$	$= BFXY$	T
ABC	$= DEF$	$= X$	$= ABCDEFX$	$= ABDEY$	$= CFY$	$= CDEXY$	$= ABFXY$	B
D	$= ABCEF$	$= ABCDX$	$= EFX$	$= CEY$	$= ABDFY$	$= ABEXY$	$= CDFXY$	T
AD	$= BCEF$	$= BCDX$	$= AEFX$	$= ACEY$	$= BDFY$	$= BEXY$	$= ACDFXY$	T
BD	$= ACEF$	$= ACDX$	$= BEFX$	$= BCEY$	$= ADFY$	$= AEXY$	$= BCDFXY$	T
ABD	$= CEF$	$= CDX$	$= ABEFX$	$= ABCEY$	$= DFY$	$= EXY$	$= ABCDFXY$	E
CD	$= ABEF$	$= ABDX$	$= CEFX$	$= EY$	$= ABCDFY$	$= ABCEXY$	$= DFXY$	T
ACD	$= BEF$	$= BDX$	$= ACEFX$	$= AEY$	$= BCDFY$	$= BCEXY$	$= ADFXY$	E
BCD	$= AEF$	$= ADX$	$= BCEFX$	$= BEY$	$= ACDFY$	$= ACEXY$	$= BDFXY$	E
$ABCD$	$= EF$	$= DX$	$= ABCEFX$	$= ABEY$	$= CDFY$	$= CEXY$	$= ABDFXY$	T
E	$= ABCDF$	$= ABCEX$	$= DFX$	$= CDY$	$= ABEFY$	$= ABDXY$	$= CEFXY$	T
AE	$= BCDF$	$= BCEX$	$= ADFX$	$= ACDY$	$= BEFY$	$= BDXY$	$= ACEFXY$	T
BE	$= ACDF$	$= ACEX$	$= BDFX$	$= BCDY$	$= AEFY$	$= ADXY$	$= BCEFXY$	T
ABE	$= CDF$	$= CEX$	$= ABDFX$	$= ABCDY$	$= EFY$	$= DXY$	$= ABCEFXY$	E
CE	$= ABDF$	$= ABEX$	$= CDFX$	$= DY$	$= ABCEFY$	$= ABCDXY$	$= EFXY$	T
ACE	$= BDF$	$= BEX$	$= ACDFX$	$= ADY$	$= BCEFY$	$= BCDXY$	$= AEFXY$	E
BCE	$= ADF$	$= AEX$	$= BCDFX$	$= BDY$	$= ACEFY$	$= ACDXY$	$= BEFXY$	E
$ABCE$	$= DF$	$= EX$	$= ABCDFX$	$= ABDY$	$= CEFY$	$= CDXY$	$= ABEFXY$	T
DE	$= ABCF$	$= ABCDEX$	$= FX$	$= CY$	$= ABDEFY$	$= ABXY$	$= CDEFXY$	T
ADE	$= BCF$	$= BCDEX$	$= AFX$	$= ACY$	$= BDEFY$	$= BXY$	$= ACDEFXY$	E
BDE	$= ACF$	$= ACDEX$	$= BFX$	$= BCY$	$= ADEFY$	$= AXY$	$= BCDEFXY$	E
$ABDE$	$= CF$	$= CDEX$	$= ABFX$	$= ABCY$	$= DEFY$	$= XY$	$= ABCDEFXY$	B
CDE	$= ABF$	$= ABDEX$	$= CFX$	$= Y$	$= ABCDEFY$	$= ABCXY$	$= DEFXY$	B
$ACDE$	$= BF$	$= BDEX$	$= ACFX$	$= AY$	$= BCDEFY$	$= BCXY$	$= ADEFXY$	T
$BCDE$	$= AF$	$= ADEX$	$= BCFX$	$= BY$	$= ACDEFY$	$= ACXY$	$= BDEFXY$	T
$ABCDE$	$= F$	$= DEX$	$= ABCFX$	$= ABY$	$= CDEFY$	$= CXY$	$= ABDEFXY$	T

The partition of the degrees of freedom in the analysis of variance which would generally be made is the following:

		D.F.
Blocks		3
Treatments:	Main effects	6
	Interactions	14
Error		8
		31

The table of aliases is condensed below by the omission of all aliases involving more than two factors—counting, as before, XY as a single factor as well as X and Y.

Effects A, B, D, E have aliases of at least three letters, but $C = FXY$ and $F = CXY$.

Effects AD, BD, AE, BE have aliases of at least three letters, but

$$AB = CX = FY, \quad AC = BX, \quad BC = AX, \quad CD = EY, \quad EF = DX,$$

$$CE = DY, \quad DE = FX = CY, \quad DF = EX, \quad BF = AY, \quad AF = BY.$$

In an experiment in which block-treatment interactions cannot be assumed to be negligible in relation to the effects it is desired to estimate, the interpretation of most two-factor interactions is difficult if not impossible. The following identities of practical interest exist for the terms which would be used to estimate the error: ACD, BCD, ACE, BCE have aliases of three letters and are either three-factor interactions or interactions between blocks and two-factor interactions, but $ABD = EXY$, $ABE = DXY$, $ADE = BXY$, and $BDE = AXY$.

This design is very similar in result to the fully replicated but confounded 2^5 design described above.

Fractional replication in the 3^n system

Here we have to consider treatment effects assessed from powers of one-third of a complete replicate. Only those treatment combinations represented by points of the lattice lying on the hyperplane

$$\alpha_1 y_1+\alpha_2 y_2+\ldots+\alpha_n y_n = 0, \text{ or } 1, \text{ or } 2 \pmod 3$$

will be included in a one-third replicate.

A particular treatment effect is given by the differences between the means of those plots represented by points on the following three planes:

$$\beta_1 y_1+\beta_2 y_2+\ldots+\beta_n y_n = 0 \pmod 3,$$
$$\beta_1 y_1+\beta_2 y_2+\ldots+\beta_n y_n = 1 \pmod 3,$$
$$\beta_1 y_1+\beta_2 y_2+\ldots+\beta_n y_n = 2 \pmod 3.$$

It is obvious that the points lying on the first plane will also lie on the planes

$$(\beta_1+\lambda\alpha_1)\, y_1+(\beta_2+\lambda\alpha_2)\, y_2+\ldots+(\beta_n+\lambda\alpha_n) y_n = 0 \pmod 3, \quad \text{for } \lambda = 1 \text{ and } 2;$$

the points on the other two planes will lie on these planes with 1 and 2 respectively on the right-hand side of the equation.

The aliases of each pair of degrees of freedom are therefore obtained by multiplication of its symbol by

$$y_1^{\alpha_1} y_2^{\alpha_2} \ldots y_n^{\alpha_n},$$

and by its square.

As an example, suppose a third replicate of a 3^3 design is based on the inclusion only of those treatment combinations represented by the symbol $y_1y_2y_3 (y_1+y_2+y_3 = 0$ say), then the aliases are exemplified by the relationship $y_1 = y_1y_2^2y_3^2 = y_2y_3$.

The confounding of one replicate of a 3^3 experiment in three blocks of nine plots

A frequently used design is the 3^3 in three blocks of nine plots, testing all combinations of three factors each at three levels. This design is formally a one-third replicate of a 3^4 design. Suppose the factors are y_1, y_2, and y_3 and let blocks be denoted by the pseudo-factor b; a three-factor interaction of y_1, y_2, and y_3, say $y_1y_2y_3$, is usually confounded in order to keep main effects and first-order interactions free of block effects.

Then $b = y_1y_2y_3$ or $I = y_1y_2y_3b^2$, since $b^3 = 1$.

As in the case of the 2^5 design, we work out the aliases of each pair of degrees of freedom: each pair of degrees of freedom will in this case have two aliases:

$$\begin{array}{llll}
y_1 = y_1y_2^2y_3^2b & = y_2y_3b^2 & \qquad y_2y_3 = y_1y_2^2y_3^2b^2 & = y_1b^2 \\
y_2 = y_1y_2^2y_3b_2^2 & = y_1y_3b^2 & \qquad y_2y_3^2 = y_1y_2^2b^2 & = y_1y_3^2b^2 \\
y_3 = y_1y_2y_3^2b^2 & = y_1y_2b^2 & \qquad y_1y_2y_3 = y_1y_2y_3b & = b^2 \\
y_1y_2 = y_1y_2y_3^2b & = y_3b^2 & \qquad y_1y_2y_3^2 = y_1y_2b & = y_3b \\
y_1y_2^2 = y_1y_3^2b & = y_2y_3^2b & \qquad y_1y_2^2y_3 = y_1y_3b & = y_2b \\
y_1y_3 = y_1y_2^2y_3b & = y_2b^2 & \qquad y_1y_2^2y_3^2 = y_1b & = y_2y_3b \\
y_1y_3^2 = y_1y_2^2b & = y_2y_3^2b^2 & &
\end{array}$$

Here again the identities could result in difficulty in interpretation—as of course could have been predicted from the examination of the possible arrangements in blocks of nine of the 3^4 design. The main effects may be regarded as clear, and three of the first-order interactions. The remaining two-factor interactions could be ascribed to differential effects of the factors on the three blocks. The three-factor interactions which are not confounded with blocks are also ascribable to interactions of main effects and blocks and may therefore be used to form an estimate of the error of these effects.

GENERAL REMARKS ON CONFOUNDING

The device of confounding is used almost without exception in agricultural experiments in order to reduce the block size to twelve or less plots. As the above results indicate there are two aspects which then need careful consideration, (*a*) the estimation of interactions, and (*b*) the estimation of the experimental error.

The main purpose of the factorial design is the estimation of main effects and interactions between pairs of factors and thence of the effect of any one factor in the presence and absence of each of the other factors. It is clear that when it is necessary to remove soil heterogeneity by confounding, the interpretation of a small experiment involving a few factors may be exceedingly difficult because of the possibility of block-treatment interactions. It is possible to use the rule that a large contrast should be regarded as the interaction between whichever pair of main effects is the larger, but this rule will break down in some cases when, for example, the contrast has two aliases AB and CD, and effects A and C are large and B and D small. In the case of a series of experiments, a device which might be helpful is the use of permutations of the possible identity relationships, one at each centre. The modern emphasis in agricultural experimentation is on series of experiments at various places and in several years, rather than on individual experiments. Interactions of pairs of factors will be estimated correctly from a large series of experiments if treatments are assigned at random to blocks.

The evaluation of two-factor interactions for individual experiments depends on the assumption that block-treatment interactions are small compared with the experimental error. Yates (1935) examined several experiments for the existence of such interactions and found no evidence of them. Since that time a large number of experimental results which can be used to provide information on the question have been accumulated, and an investigation of these has indicated that block-treatment interactions are negligible and may be ignored (Kempthorne, 1947).

With regard to the estimation of error, in so far as tests of significance are of interest, it can be said that the analysis of variance does provide a test of significance of the hypothesis that the treatments have an overall effect different from zero. In agricultural experimentation, the term error is used to denote block-treatment interactions. Thus in the simple randomized block experiment, it is possible to evaluate the difference between two treatments from each block, and it is the variability of this difference from block to block which is regarded as the error. In general, as there are usually few blocks, and the error of each comparison would be determined with poor accuracy, the errors of all the possible independent comparisons are pooled to give a common estimate. If the treatments were duplicated at random

within each block, the analysis would be of the form (r being the number of blocks and t of treatments):

	D.F.
Blocks	$r-1$
Treatments	$t-1$
Treatments by blocks	$(r-1)(t-1)$
Within blocks	rt
	$2rt-1$

The component 'within blocks' could more accurately be described as experimental error, but would not be used to evaluate the errors of treatment effects, since the experimenter is interested in the constancy of treatment effects from block to block. There is therefore little point in actually carrying out such an experiment. In a factorial experiment with replication, the components which could be evaluated consist of replicates, effects and low-order interactions, high-order interactions, and interactions of treatments and replicates. On the assumption that the sum of squares for interaction of treatments and replicates is homogeneous, the mean square for high-order interactions will include the mean square for treatments × replicates plus a component of variance due to high-order interactions. When only one replication is used, it is assumed that the component of variance due to high-order interactions is small, and that the high-order interactions mean square can be regarded as an estimate of error. It is important to bear in mind that an individual agricultural experiment can give information only for a particular set of experimental conditions and that it is known from experience that place to place and year to year variability is considerable. It would therefore be uneconomical to utilize available resources to determine effects and their errors at a few particular places very accurately, but preferable to sacrifice replication at each place in order to have information over a large range of experimental conditions.

Mixed systems

It is not proposed to examine mixed systems of the type $p^m q^n$, where p and q are primes, in the present paper. It is clear, however, that the possibilities of complete confounding and fractional replication are very limited. A p'th replicate must obviously include p^{m-1} combinations of the m factors combined with all the q^n combinations of the n factors. For the examination of treatment aliases the system may be regarded as the product of the two separate systems. Thus if $p = 3$, $m = 2$, $q = 2$, $n = 3$ and the factors are $y_1 y_2 y_3' y_4' y_5'$, then a half replicate would be obtained by putting $I = y_3' y_4' y_5'$. The aliases which result are exemplified by the following:

$$y_1 = y_1 y_3' y_4' y_5', \qquad y_1 y_2 y_3' = y_1 y_2 y_4' y_5',$$
$$y_1 y_2 = y_1 y_2 y_3' y_4' y_5', \qquad y_3' = y_4' y_5'.$$

Such designs with fractional replication or complete confounding are therefore useful only when the corresponding designs for the two separate systems are feasible.

Comments on 'The design of optimum multifactorial experiments'

In a paper entitled 'The Design of Optimum Multifactorial Experiments', Plackett & Burman (1946) put forward designs more specifically for physical and industrial research, which are of interest from the point of view of fractional replication. In order to estimate the

effect of varying nine components, of an assembly, each component having two possible values, a nominal (−) and an extreme (+), they put forward the following design which requires the testing of sixteen assemblies:

	Components								
	1	2	3	4	5	6	7	8	9
Assembly 1	+	−	−	−	+	−	−	+	+
2	+	+	−	−	−	+	−	−	+
3	+	+	+	−	−	−	+	−	−
4	+	+	+	+	−	−	−	+	−
5	−	+	+	+	+	−	−	−	+
6	+	−	+	+	+	+	−	−	−
7	−	+	−	+	+	+	+	−	−
8	+	−	+	−	+	+	+	+	−
9	+	+	−	+	−	+	+	+	+
10	−	+	+	−	+	−	+	+	+
11	−	−	+	+	−	+	−	+	+
12	+	−	−	+	+	−	+	−	+
13	−	+	−	−	+	+	−	+	−
14	−	−	+	−	−	+	+	−	+
15	−	−	−	+	−	−	+	+	−
16	−	−	−	−	−	−	−	−	−

Yates put forward a similar design in his 1935 paper for the weighing of a number of small articles on a balance which required a zero correction, as an example of the estimation of the effects of independent factors. In his case there was a close formal analogy to the 2^n factorial system, and it will now be shown that Plackett & Burman's design given above is a high-order fractional design of the type discussed in the present paper.

Denoting the nominal values by unity and the extreme values of the nine components by $a, b, c, d, e, f, g, h, k$ in order, the treatment combinations represented are 1, $aehi$, $abfi$, $abcg$, $abcdh$, $bcdei$, $acdef$, $bdefg$, $acefgh$, $abdfghi$, $bceghi$, $cdfhi$, $adegi$, $befh$, $cfgi$, dgh. It is found merely by one-by-one examination of the three-factor interactions that all the above sets of treatment combinations occur with the same sign in the following:

$$ABE, \quad ACK, \quad BCF, \quad CDG, \quad DEH.$$

The same will be true for all the members of the Abelian group of which the above five interactions are generators. The identity relationship is therefore:

$$\begin{array}{llllllll}
I & = ABE & = ACK & = BCEK & = BCF & = ACEF & = ABFK & = EFK \\
= CDG & = ABCDEG & = ADGK & = BDEGK & = BDFG & = ADEFG & = ABCDFGK & = CDEFGK \\
= DEH & = ABDH & = ACDEHK & = BCDHK & = BCDEGH & = ACDFH & = ABDEFHK & = DFHK \\
= CEGH & = ABCGH & = AEGHK & = BGHK & = BEFGH & = AFGH & = ABCEFGHK & = CFGHK
\end{array}$$

The identities of interest to the experimenter are the following:

$$I = ABE = ACK = BCF = EFK = CDG = DEH;$$

from these we derive the following aliases for main effects:

$$\begin{array}{ll}
A = BE = CK, & F = BC = EK, \\
B = AE = CF, & G = CD, \\
C = AK = BF = DG, & H = DE, \\
D = CG = EH, & K = AC = EF. \\
E = AB = FK = DH, &
\end{array}$$

In all cases, the contrasts estimating main effects are minus the contrasts estimating interactions. If, for example, the interaction of B and E is negative, and A has no effect, the

conclusion drawn by the experimenter will be that A has a positive effect. It is possible but rather difficult to imagine physical systems in which effects will not interact, and interpretation of the results of experiments based on this design may often be impossible. With nine factors, it appears from the present work that the minimum number of combinations which should be tested is 128, that is one-quarter of a replication, though it is possible that by making less stringent assumptions about two-factor interactions, one-eighth of a replication might give intelligible results. A possible instance in which it might be feasible to use the designs discussed is when it is expected that only one or two of the factors have an effect, and the problem is to determine as quickly as possible which of the nine factors are responsible. An example in which a high-order fractional design was used in such circumstances with good results has been described by Tippett (1936). A detailed examination of all the designs put forward by Plackett & Burman will not be undertaken, but the lines on which such an examination would proceed and the broad conclusions which would emerge are obvious from the above examination of one of their simpler designs.

Conclusions

A method of examining fractional replication and confounding for some types of factorial experiments is described. The formal equivalence between the two is indicated and the implications of this equivalence discussed. Further progress will follow on group theory lines and this is being examined, together with the possibility of fractional replication when the fraction is greater than unity. The possibilities are explored of the estimation of main effects and two-factor interactions of many factors by testing only a small proportion of the possible treatment combinations. An examination on these lines is made of designs proposed by Plackett & Burman.

References

Finney, D. J. (1945). The fractional replication of factorial arrangements. *Ann. Eugen., Lond.*, **12**, 291–301.

Fisher, R. A. (1942). The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugen., Lond.*, **11**, 341–53.

Kempthorne, O. (1947). A note on differential responses in blocks. *J. Agric. Sci.* **37**, 245–48.

Plackett, R. L. & Burman, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika*, **33**, 305–25.

Tippett, L. H. C. (1936). *Applications of Statistical Methods to the Control of Quality in Industrial Production.* Manchester Statist. Soc.

Yates, F. (1935). Complex experiments. *J. R. Statist. Soc. Suppl.* **2**, 181–223.

Yates, F. (1937). The design and analysis of factorial experiments. *Tech. Commun. Imp. Bur. Soil Sci.*, no. 35.

A COMPARISON OF STRATIFIED WITH UNRESTRICTED RANDOM SAMPLING FROM A FINITE POPULATION*

By P. ARMITAGE, B.A.

1. Introduction

1·1. We are concerned in this paper with the problem of estimating the mean value μ of a variable x in a population, by taking a sample which is in some way representative of the population. It has been realized since Bowley's paper (1926), and more particularly since Neyman's more comprehensive survey (1934), that a certain degree of precision in the estimate can often be obtained more economically by stratified random sampling (usually referred to merely as *stratified sampling*) than by unrestricted random sampling (usually called merely *random sampling*). In the stratified method, the population is divided into several strata, the sample size divided in some prearranged way among the strata, and sampling performed at random from each stratum. In unrestricted random sampling, a random selection is made from the whole population, and the method may be regarded as a particular case of stratification, where the number of strata is one.

Some text-books deal briefly with stratified sampling. Wilks (1943) considers only infinite populations, and denotes by representative sampling what we should call a particular type of stratified sampling (see § 1·2). The subject is treated by Kendall (1946, pp. 249–52), but he makes no comparison with unrestricted random sampling. We shall begin by introducing several well-known results which will be needed later.

1·2. The summation sign $\sum$ will be used throughout for $\sum\limits_{l=1}^{r}$, and $\sum\limits_{k}$ for $\sum\limits_{k=1}^{r}$. In general, $\sum$ is used for a single summation, $\sum\limits_{k}\sum$ for a double summation, and the suffix k where no summation is involved.

We shall consider the following position: A population π of size N is subdivided into r strata, π_k, of size N_k ($\sum N_l = N$). The variable x is distributed so that the mean and variance (divisor N_k) within π_k are respectively μ_k, σ_k^2. It is required to estimate $\mu = \sum N_l\mu_l/N$, the grand mean.

Suppose a given sample size, n, is divided so that n_k items are sampled at random from π_k ($\sum n_l = n$). We may denote the jth observation from the kth sample by x_{kj} ($j = 1, 2, \ldots, n_k$), and the mean and variance of the kth sample by $\bar{x}_k$ and s_k^2, which are known to be unbiased estimates of μ_k and $\dfrac{(n_k-1)\,N_k\sigma_k^2}{n_k(N_k-1)}$, respectively (see, for example, Kendall, 1943, p. 284). It seems intuitively obvious to take as our estimate of μ,

$$m = \sum N_l\bar{x}_l/N, \tag{i}$$

which is clearly unbiased. This is, however, not the only unbiased estimate which is a linear function of the x_{kj}. For instance, $\sum N_l x_{l1}/N$ also satisfies the conditions. Neyman (1934) has shown that, for fixed values of n_k, the estimate given by (i) is the best linear unbiased estimate of μ, in the sense that its sampling variance is less than that of any other linear unbiased estimate.

* Communication from the National Physical Laboratory.

The question now arises: given a sample size n, how shall we choose the n_k so as to minimize var (m), where m is given by (i)? Bowley had not considered 'best' estimates, and he suggested that n_k should be proportional to N_k, i.e.

$$n_k = \frac{nN_k}{N}. \tag{ii}$$

Neyman (1934) showed, by the method given in § 2, that the values of n_k which minimize var (m) are

$$\begin{aligned} n_k &= \frac{nN_k\sigma_k\sqrt{[N_k/(N_k-1)]}}{\sum N_l\sigma_l\sqrt{[N_l/(N_l-1)]}} \\ &= \frac{nN_k\sigma'_k}{\sum N_l\sigma'_l}, \end{aligned} \tag{iii}$$

where $\sigma'_k = \sigma_k\sqrt{[N_k/(N_k-1)]}$.

We shall refer to these two methods of defining the n_k, by (ii) and (iii) respectively, as *proportionate sampling*, and *optimum stratified sampling*, denoting by m_p and m_o the estimates of μ obtained from (i) by the two methods, and by $\bar{x}$ the estimate of μ given by the mean of an unrestricted random sample of n from the whole population π.

The optimum stratified method thus requires a knowledge of the σ_k. In practice, we should never know the σ_k exactly, unless the population had been subjected to exhaustive sampling, in which case μ would be known exactly. Sukhatme (1935) has shown that, at any rate for large N_k, if the σ_k^2 are estimated from a preliminary sample, and the n_k defined by using these estimates in (iii), there is a high probability that var $(m_o) <$ var (m_p).* The efficiency of this method will of course depend on the size of the preliminary sample, and Sukhatme's investigation only dealt with one value of this (15 from each stratum). In some cases we should be able to form a fairly good estimate of the σ_k from past experience, and there would be no need for a preliminary sample.

Another interesting comparison which has not been extensively investigated is that between optimum stratified sampling and unrestricted random sampling. Wilks (1943) deals with this for infinite populations, and obtains (pp. 88, 89) the result (in our notation),

$$\text{var}\,(m_o) \leqslant \text{var}\,(m_p) \leqslant \text{var}\,(\bar{x}), \tag{iv}$$

the first equality holding only when all the σ_k are equal, and the second only when all the μ_k are equal. (Our N_k/N are replaced by p_k, where p_k is the probability that x, when drawn at random from π, is a member of π_k, so that, for instance, (iii) becomes

$$n_k = \frac{np_k\sigma_k}{\sum p_l\sigma_l}.$$

Representative sampling as defined by Wilks is what we should call proportionate sampling.) We shall show in § 2 that for finite populations, while the relation

$$\text{var}\,(m_o) \leqslant \text{var}\,(m_p) \tag{v}$$

is always true, the equality holding only when all the σ'_k are equal, it is not necessarily true that

$$\text{var}\,(m_p) \leqslant \text{var}\,(\bar{x}), \tag{vi}$$

* No confusion need arise from the fact that the symbol m_o and the term *optimum* are still used when estimates of the σ_k are used in (iii).

and in fact in the limiting case when all the μ_k are equal, it is true that

$$\text{var}(m_p) > \text{var}(\bar{x}), \tag{via}$$

so that if the σ'_k are also equal

$$\text{var}(m_0) > \text{var}(\bar{x}); \tag{vib}$$

i.e. random sampling gives a more accurate estimate of the mean than any stratified sampling. We shall see, however, that in almost all practical cases (iv) is true.

2. Derivation of formulae

2·1. *Results* (iii) *and* (v). Using the notation of § 1·2, we have the standard result that

$$\text{var}(\bar{x}_k) = \frac{\sigma_k^2}{n_k}\left(\frac{N_k - n_k}{N_k - 1}\right) \quad \text{(see e.g. Wilks, p. 86).} \tag{vii}$$

Therefore from (i),

$$\begin{aligned}\text{var}(m) &= \sum \frac{N_l^2 \sigma_l^2}{N^2 n_l}\left(\frac{N_l - n_l}{N_l - 1}\right) \\ &= \sum \frac{N_l \sigma_l'^2}{N^2 n_l}(N_l - n_l).\end{aligned} \tag{viii}$$

The result (iii) may be obtained quite easily by finding the values of the n_l which minimize (viii) subject to the condition $\sum n_l = n$, using the method of Lagrange multipliers. Then, substituting (ii) and (iii) in (viii), and applying Schwarz's inequality, we have (v). The following method is due to Neyman.

It may be verified from (viii) that

$$\text{var}(m) = \frac{N-n}{N^2 n}\sum N_l \sigma_l'^2 + \frac{1}{N^2}\sum_k n_k\left(\frac{N_k\sigma'_k}{n_k} - \frac{\sum N_l \sigma'_l}{n}\right)^2 - \frac{1}{Nn}\sum_k N_k\left(\sigma'_k - \frac{\sum N_l\sigma'_l}{N}\right)^2. \tag{ix}$$

If we denote the three terms of (ix) by A, B and C, so that

$$\text{var}(m) = A + B - C,$$

it will be seen that A and C are independent of n_k and, since B is non-negative, it follows that the values of n_k which minimize var (m) must minimize B. Now $B = 0$ if and only if

$$n_k = \frac{nN_k\sigma'_k}{\sum N_l \sigma'_l},$$

which is (iii). For these values of n_k, $m = m_0$, and

$$\text{var}(m_0) = A - C. \tag{x}$$

If we define n_k by (ii), so that $m = m_p$, we see from (ix) that $B = C$, so that

$$\text{var}(m_p) = A. \tag{xi}$$

From (x) and (xi), we obtain (v), the equality holding only when $C = 0$, which is true only when $\sigma'_k - \dfrac{\sum N_l \sigma'_l}{N} = 0$ for all k, i.e. when the σ'_k are all equal.

2·2. *Unrestricted random sampling.* The variance of a random observation x from π is

$$\sigma^2 = \frac{\sum N_l \sigma_l^2}{N} + S,$$

where S is the weighted sum of squares of the μ_l, i.e. $S = \frac{\sum N_l(\mu_l - \mu)^2}{N}$. From (vii),

$$\begin{aligned}\operatorname{var}(\bar{x}) &= \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)\\ &= \frac{N-n}{nN(N-1)}\sum N_l\sigma_l^2 + \frac{N-n}{n(N-1)}S\\ &= \frac{N-n}{nN(N-1)}\sum(N_l-1)\sigma_l'^2 + \frac{N-n}{n(N-1)}S.\end{aligned}$$

From (xi),
$$\operatorname{var}(m_p) = \frac{N-n}{N^2 n}\sum N_l\sigma_l'^2.$$

Denoting $\frac{\sum(N_l-1)\sigma_l'^2}{N-1}$ by H, and $\frac{\sum N_l\sigma_l'^2}{N}$ by K, we have

$$\left.\begin{aligned}\operatorname{var}(\bar{x}) &= \frac{N-n}{Nn}H + \frac{N-n}{(N-1)n}S\\ \text{and}\qquad \operatorname{var}(m_p) &= \frac{N-n}{Nn}K.\end{aligned}\right\} \tag{xii}$$

Now
$$H-K = \frac{\sum N_l\sigma_l'^2 - \sum N_l\sum\sigma_l'^2}{N(N-1)} < 0,$$

and if we regard each N_l as being of the same order, $O(N)$, then $H-K$ is $O(N^{-1})$, which means that when all the μ_k are equal, $S=0$, and so

$$\operatorname{var}(\bar{x}) < \operatorname{var}(m_p),$$

which is (vi$a$); but as $N \to \infty$,
$$\operatorname{var}(\bar{x}) \sim \operatorname{var}(m_p) + S/n, \tag{xiii}$$

giving Wilks's result (p. 88) that for infinite populations (vi) is true, the equality holding only when all the μ_k are equal.

From (v) and (vi$a$) it follows that for finite populations, when all the μ_k are equal and all the σ_k' are equal, (vi$b$) is true, i.e. in this case unrestricted random sampling is actually better than any stratified random sampling with the same sample size.

3. General comparison

3·1. From (ix), (x) and (xii),

$$\phi \equiv \operatorname{var}(\bar{x}) - \operatorname{var}(m_0) - \frac{N-n}{(N-1)n}S = \frac{1}{N^2(N-1)n}(P-Q-R), \tag{xiv}$$

where
$$\begin{aligned}P &= N^2\sum N_l\sigma_l'^2 - N(\sum N_l\sigma_l')^2 \geqslant 0 \quad \text{(equality if all } \sigma_l' \text{ are equal)},\\ Q &= n(\sum N_l\sigma_l'^2 - N\sum\sigma_l'^2) < 0,\\ R &= N^2\sum\sigma_l'^2 - (\sum N_l\sigma_l')^2 > 0.\end{aligned}$$

As $N \to \infty$, P, Q and R are respectively $O(N^3)$, $O(N)$ and $O(N^2)$, and so we have the result that for infinite populations $\phi \geqslant 0$, which with (xiii) is easily seen to be equivalent to Wilks's result (iv).

In the finite case, however, by suitable choice of the σ_l' and n we can make ϕ either positive or negative. For instance, if the σ_l' are all equal and n is sufficiently small, R predominates in (xiv), and $\phi < 0$. As n increases to N, ϕ increases to 0. (By considering Q and R, it is not

obvious that $\phi \to 0$ in this case, but it must be remembered that (xiv) is only true if the n_k are given by (iii), and this becomes impossible as n approaches N. This will be remarked upon below.) If the σ'_l are sufficiently unequal, P will predominate and $\phi > 0$. In this case the factor $\frac{(P-R)}{N^2(N-1)n}$ in (xiv) will be positive, and ϕ will decrease as n increases.

The situations, then, in which (vi$b$) is likely to be true (provided that the n_k are really given by (iii)) are when the μ_k are nearly equal, and when N is small or the σ'_k are nearly equal. We shall consider some examples in § 4.

3·2. In applying the procedure of stratification, we shall make two departures from the theory outlined above which will tend to nullify the advantages of the stratified method. The first is that, as was pointed out in § 1·2, we shall never know the σ_k exactly, and the degree to which our estimates from which the n_k were obtained are accurate depends on the circumstances. It seems quite likely that Sukhatme's result will be fairly well applicable to finite populations, but there is an opportunity for research on this point.

The second respect in which we depart from theory lies in the fact that, even if the σ_k are exactly known, the n_k that we choose can never be exactly as given by (iii); first because they must be integers, which makes a considerable difference when n is small (the size of the smallest stratified sample from which an unbiased estimate of μ can be made is clearly r); and secondly, n_k cannot take values greater than N_k. In this latter case, if the values of, say, s of the n_k, as given by (iii), are greater than the corresponding N_k, we should let $n_k = N_k$ for these s strata, and then set the other $(r-s)$ values of n_k proportional to the corresponding $N_k\sigma'_k$. This will clearly decrease $\text{var}(\bar{x}) - \text{var}(m_o)$ as given by (xiv). For example, when $n = N$, we have

$$\text{var}(\bar{x}) = \text{var}(m_o) = \frac{(N-n)}{(N-1)n} S = 0,$$

but the right-hand side of (xiv)

$$= \frac{1}{N^3}\{N\Sigma N_l \sigma_l'^2 - (\Sigma N_l \sigma'_l)^2\} \geqslant 0$$

(equality holding if all the σ'_l are equal). In fact both these limitations will decrease the theoretical advantage (if any) of stratified over random sampling, and we must take them into account in assessing the relative merits of the two methods.

4. Examples

In the four examples illustrated by Figs. 1–4, $\text{var}(m_o)$ and $\text{var}(\bar{x})$ have been calculated for different stratified populations, and $\psi = \log_{10}\{\text{var}(\bar{x})/\text{var}(m_o)\}$ plotted against $c = n/N$, so that $\psi < 0$ if $\text{var}(\bar{x}) < \text{var}(m_o)$. In each figure the different curves represent populations with the same σ_k, with the N_k in the same proportions but with different magnitudes, and with the μ_k equal, so that $S = 0$.

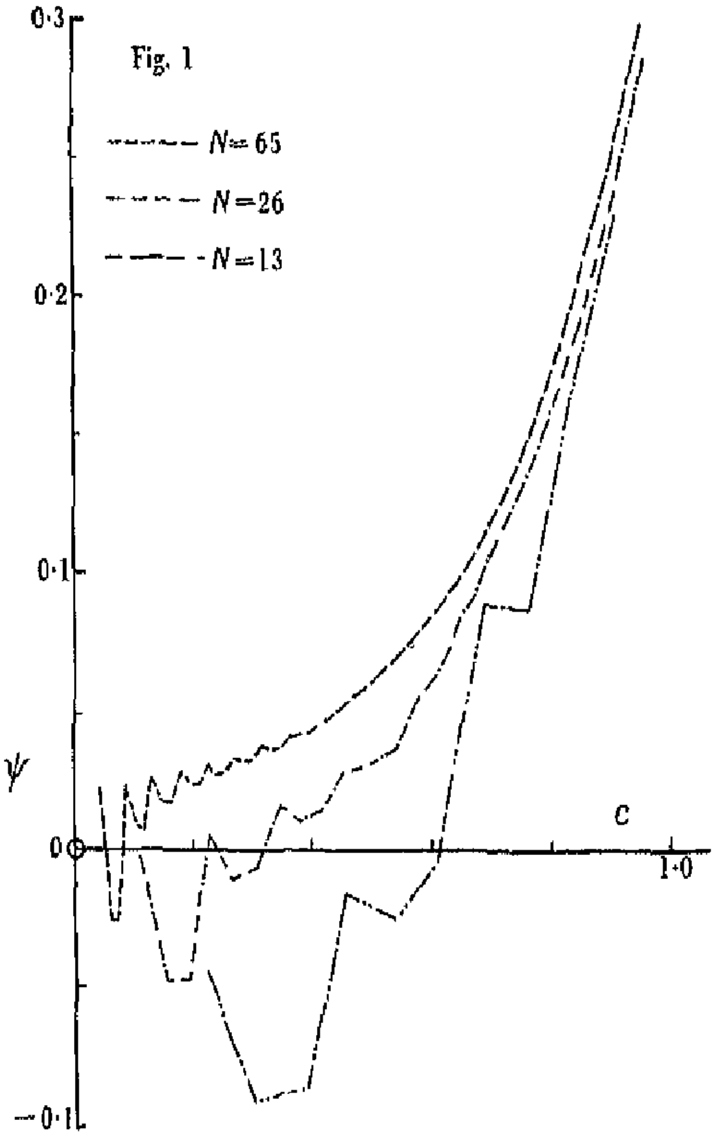
Example 1. $\sigma_k = 2, 3, 4,$ $N_k \propto 6, 4, 3$ $(N = 65, 26, 13)$.

Example 2. $\sigma_k = 4, 5, 6,$ $N_k \propto 6, 5, 4$ $(N = 120, 60, 30, 15)$.

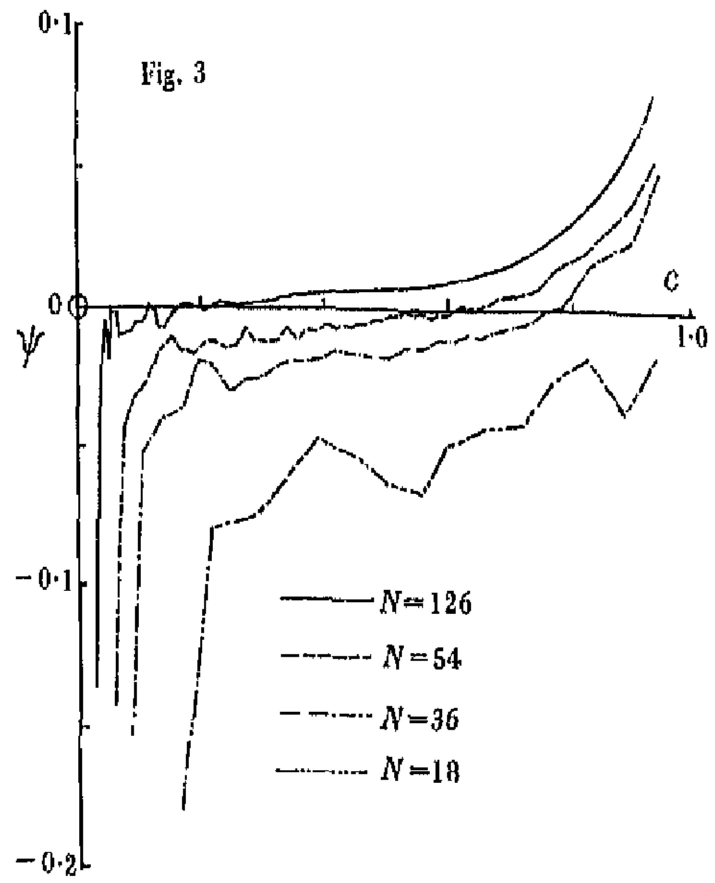
Example 3. $\sigma_k = 4, 5, 6,$ $N_k \propto 3, 11, 4$ $(N = 126, 54, 36, 18)$.

Example 4. $\sigma_k = 1, 1, 2, 3, 4,$ $N_k \propto 5, 5, 1, 2, 3$ $(N = 128, 32)$.

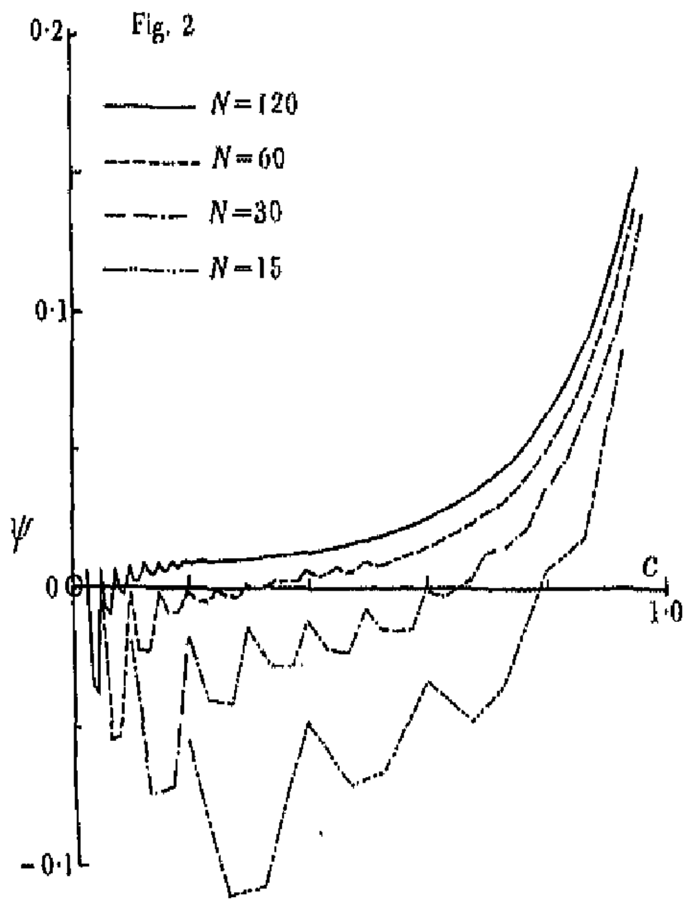
The first thing to be noticed about the graphs is that in each one ψ increases, generally speaking, as n increases. Further, in any one example the range of c for which $\psi < 0$ increases



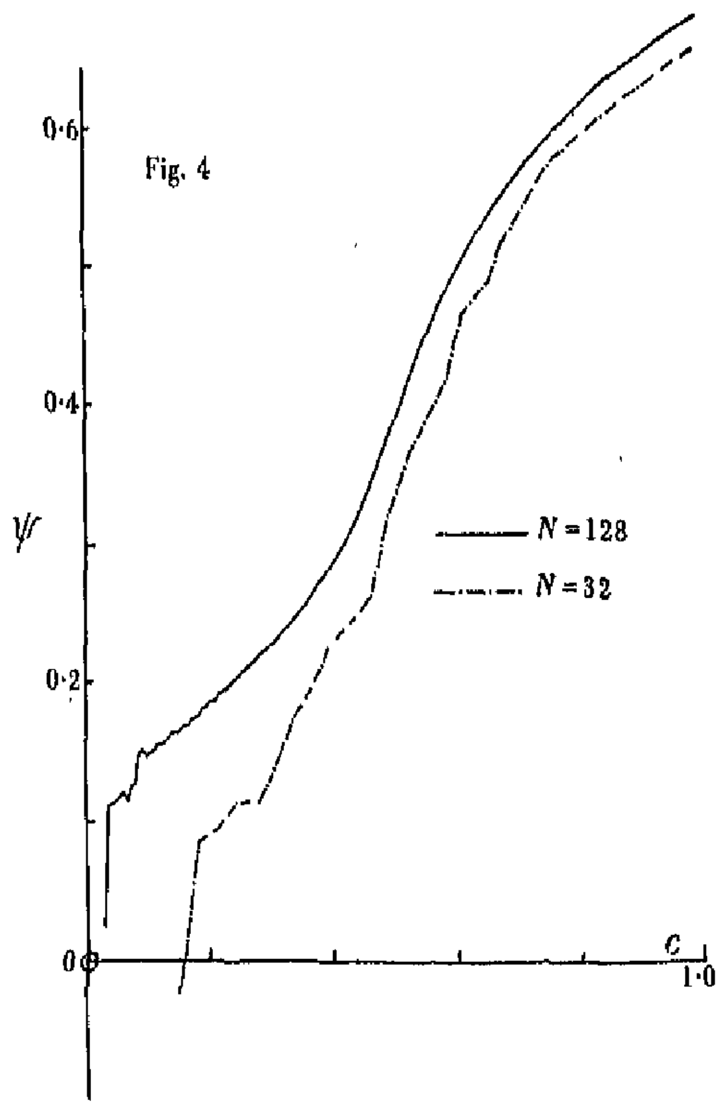
0·3
Fig. 1
N=65
N=26
N=13
0·2
0·1
ψ
0
c
1·0
−0·1



0·1
Fig. 3
0
c
ψ
1·0
−0·1
N=126
N=54
N=36
N=18
−0·2



0·2
Fig. 2
N=120
N=60
N=30
N=15
0·1
ψ
0
c
1·0
−0·1



0·6
Fig. 4
0·4
ψ
N=128
N=32
0·2
0
c
1·0

as N decreases; and in this sense we can say that for small samples of proportionate size from a stratified population, the advantage (if any) of the stratified method decreases as N decreases.

Secondly, the curves are not smooth. The reason for this is clear. In the optimum stratified method the n_k are to be chosen approximately proportional to $N_k\sigma_k$ (a second approximation is $(N_k+\frac{1}{2})\sigma_k$). In Example 1, the $N_k\sigma_k$ are all equal, and it follows that the n_k should be nearly equal. If $n\equiv 0 \pmod 3$ this can be done, but for $n\equiv 1, 2 \pmod 3$, $\mathrm{var}\,(m_o)$ takes values greater than it would if fractional n_k were allowed. This produces a rise in the curve of ψ for $n\equiv 0 \pmod 3$, which gradually disappears as n increases since the effect is much greater for small n. The same 'period' is noticeable in Fig. 2, but in Figs. 3 and 4, where the main 'periods' are respectively 15 and 30, the effect is smaller.

We saw in § 3·1 that, broadly speaking, the advantage of the stratified method decreases as the σ_k tend to equality. This is illustrated by comparing Examples 1 and 2. In each of these the $N_k\sigma_k$ are equal, but in Example 2 the σ_k are proportionally more nearly equal. Comparing curves for about the same N ($N=65, 26, 13$ in Fig. 1 with $N=60, 30, 15$ in Fig. 2), we see that in Fig. 2 the range of values of c for which $\psi<0$ is greater than in Example 1.

Fig. 3 has the same σ_k as Fig. 2, but the $N_k\sigma_k$, and therefore the n_k, are different. The curves are similar to those of Fig. 2, but the stratified method is still less advantageous (especially for small values of c).

Example 4 has five instead of three strata, and there is quite large variation between the σ_k and between the $N_k\sigma_k$. There is no doubt here that $\psi>0$, the only exception being for $N=32$, $n=5$, where $\psi=-0{\cdot}02$.

These examples may be said to give the maximum advantage to the stratified method, in the sense that the calculated values of $\mathrm{var}\,(m_o)$ depend on the best method of choosing the n_k. If the σ_k are not sufficiently well known to enable the best values of n_k to be used, then we shall get a larger value of $\mathrm{var}\,(m_o)$. It must be remembered, however, that in all these examples we assumed that there was no variation between the μ_k, a situation which would be very unlikely to occur in practice. Now it is clear from (xii) that if the same N_k and σ_k are considered as in one of the above examples, but the μ_k are now unequal, the effect is to increase the value of $\mathrm{var}\,(\bar{x})$ by $(N-n)\,S/(N-1)\,n$, where $S=\sum N_i(\mu_i-\mu)^2/N$; so, in any example where $\psi<0$ for some particular values of N and n, we can reverse the direction of the inequality by choosing a sufficiently large value of S, say

$$S_0=[\mathrm{var}\,(m_o)-\mathrm{var}\,(\bar{x})]\,(N-1)\,n/(N-n).$$

In comparing different values of S_0 for different examples, it must be remembered that the order of magnitude of S_0 depends on the σ_k and a suitable measure of comparison will be S_0/σ_0^2, where σ_0^2 is the pooled variance within strata $=\sum N_i\sigma_i^2/N$.

In Example 1, the largest value of S_0 is for $N=13$, $n=4$. Here $\mathrm{var}\,(m_o)=1{\cdot}9172$, $\mathrm{var}\,(\bar{x})=1{\cdot}5577$, and $S_0=1{\cdot}917=0{\cdot}231\sigma_0^2$. (If $\mu_1=0$, $\mu_2=2$, $\mu_3=3{\cdot}5$, then $S=2{\cdot}066$.)

In Example 2, the largest value of S_0 is for $N=15$, $n=4$. Here $\mathrm{var}\,(m_o)=24{\cdot}647$, $\mathrm{var}\,(\bar{x})=19{\cdot}119$, and $S_0=28{\cdot}14=0{\cdot}289\sigma_0^2$. (If $\mu_1=0$, $\mu_2=7$, $\mu_3=13$, then $S=28{\cdot}51$.)

In Example 3, the largest value of S_0 is for $N=18$, $n=3$. Here $\mathrm{var}\,(m_o)=46{\cdot}235$, $\mathrm{var}\,(\bar{x})=30{\cdot}523$, and $S_0=53{\cdot}42=0{\cdot}515\sigma_0^2$. (If $\mu_1=0$, $\mu_2=8$, $\mu_3=17$, then $S=57{\cdot}5$.)

In Example 4, the largest value of S_0 is for $N=32$, $n=5$ (the only occasion in this example where $\psi<0$). Here $\mathrm{var}\,(m_o)=0{\cdot}91406$, $\mathrm{var}\,(\bar{x})=0{\cdot}87097$, and $S_0=0{\cdot}2474=0{\cdot}049\sigma_0^2$. (If $\mu_1=\mu_2=0$ and $\mu_3=\mu_4=\mu_5=1$, then $S=0{\cdot}285$.)

5. Conclusions

We have seen in §3 that optimum stratified sampling may give a less accurate estimate of μ than unrestricted random sampling when the μ_k are nearly equal, and when N is small or the σ_k' are nearly equal. The examples of §4 bear out these conclusions and show that the effect is greatest for small n, Fig. 3 providing an additional suggestion that if the products $N_k\sigma_k$ are widely different the advantage of the stratified method tends to be nullified. In practice, we should probably only apply stratified sampling if we knew that the strata were sufficiently distinct to ensure considerable variation between either the μ_k or the σ_k. In the first case, if nothing much was known about the σ_k and a preliminary sample on the lines suggested by Sukhatme was impracticable, we should use proportionate sampling, and the size of S would usually ensure that $\text{var}(m_p) < \text{var}(\bar{x})$. In the second case, we should use optimum stratified sampling, and rely on the variability of the σ_k to ensure that $\text{var}(m_o) < \text{var}(\bar{x})$. Since an adequate degree of knowledge about the σ_k would be unlikely unless the N_k were quite large, we should in this case almost certainly be safe in using the method. To the above considerations must be added the fact that if very inaccurate estimates of the σ_k are used in (iii), then, whatever the nature of the population, the resulting procedure may be extremely inefficient.

It must be realized, of course, that even if it were known that $\text{var}(m_o) < \text{var}(\bar{x})$, it would not follow that the optimum stratified method would necessarily be the most convenient. It may be impossible, or at any rate inconvenient, to do any sort of random sampling, and some sort of quasi-random sampling may have to be used (see, e.g. Madow & Madow, 1944), but if the principle of random sampling is applicable the stratified method is not likely to be much more inconvenient, and in fact in most cases will be more convenient, than the unrestricted method.

Summary

The stratified method has been used in the past almost solely for large-scale social and agricultural surveys. Here the stratum sizes are large, and known results for infinite populations apply. There seems no reason why stratified sampling should not be used to advantage for smaller populations, and it is important to know to what extent these results still apply. In this paper a comparison has been made with unrestricted random sampling in the usual case where we are interested in estimating the mean. The advantages of the stratified method are modified, but in most cases where the method is applicable it will be found to be worth while.

The above work was carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory. The author desires to acknowledge the assistance rendered by Mr D. V. Lindley who prepared the diagrams.

References

Bowley, A. L. (1926). Measurement of the precision attained in sampling. *Bull. Int. Inst. Statist.* **22**, 1ère livraison.

Kendall, M. G. (1943). *The Advanced Theory of Statistics*, **1**. Griffin and Co.

Kendall, M. G. (1946). *The Advanced Theory of Statistics*, **2**. Griffin and Co.

Madow, W. G. & Madow, L. H. (1944). On the theory of systematic sampling. *Ann. Math. Statist.* **15**, 1–24.

Neyman, J. (1934). On the two different aspects of the representative method. *J. Roy. Statist. Soc.* **97**, 558–625.

Sukhatme, P. V. (1935). Contribution to the theory of the representative method. *Suppl. J. Roy. Statist. Soc.* **2**, 253–68.

Wilks, S. S. (1943). *Mathematical Statistics*. Princeton University Press.

SOME THEOREMS ON TIME SERIES. I

By P. A. P. MORAN
Institute of Statistics, Oxford University

One of the principal problems in the theory of time series is to discuss the relation between two series, and in the present paper we prove a theorem by which we can test whether two such series are independent. Such a test of significance must depend on the models which we assume for the probability processes which generate the series. In practice, the two most useful models are, first, that of a moving average of a series of independent random components and, secondly, the solutions of linear stochastic difference equations.

Let
$$\dots, \eta(t-1), \eta(t), \eta(t+1), \dots$$
be a sequence of independent random variables each distributed in the same distribution which we take to have zero mean and its second, third, and fourth moments finite. Then the time series generated by
$$X(t) = \sum_{i=0}^{N} \alpha_i \eta(t-i)$$
is a moving average with weights α_i. On the other hand, consider a stochastic difference equation of the form
$$X(t) + a_1 X(t-1) + \dots + a_h X(t-h) = \eta(t). \tag{1}$$

In order that the solution of (1) for successive values of t shall form a stationary series it is necessary to impose the condition that the roots of the characteristic equation
$$z^h + a_1 z^{h-1} + \dots + a_h = 0 \tag{2}$$
shall all lie inside the circle $|z| = 1$ (Wold, 1938, p. 53). When this is true the solution of (1) can be shown to be of the form
$$X(t) = \sum_{i=0}^{\infty} \alpha_i \eta(t-i),$$
where the α_i are certain functions of the roots of (2). In this case $\sum_{i=0}^{\infty} |\alpha_i|$ is majorized by a convergent geometric series.

Thus we see that both the above models are included in the more general one in which we define $X(t)$ as given by
$$X(t) = \sum_{i=0}^{\infty} \alpha_i \eta(t-i),$$
where the α_i are any sequence of constants satisfying $\sum_{i=0}^{\infty} |\alpha_i| < \infty$. Now suppose
$$\dots, \zeta(t-1), \zeta(t), \zeta(t+1), \dots$$
is another sequence of independent random variables having a distribution with zero mean and finite second, third and fourth moments. We write
$$Y(t) = \sum_{i=0}^{\infty} \beta_i \zeta(t-i),$$
where $\sum_{i=0}^{\infty} |\beta_i| < \infty$. To discuss whether two such empirical series of this form are correlated we prove that the covariance
$$S = \sum_{1}^{n} X(t)\, Y(t) \tag{3}$$

tends, as n increases, to be distributed in the normal form about zero mean with a second moment which is a function of the α_i and the β_i. We shall discuss later the calculation of this second moment from empirical series, in which case some care is necessary.

We first illustrate our method of proof by considering the much simpler problem of determining the asymptotic distribution of the sum

$$T_n = \sum_{s=1}^{n} X(t-s). \tag{4}$$

We shall show that this asymptotic distribution is also, under certain conditions, normal. This result is interesting because it establishes a central limit theorem (and therefore a law of large numbers) for stationary stochastic processes of this type. The law of large numbers for Markov chains has been considered by several writers, in particular Bernstein (1927), who proves his results by using central limit theorems for non-independent components. His theorems cannot be applied in the present case, but some of the ideas of his methods can.

Consider (4) above, where $X(t)$ is defined by

$$X(t) = \sum_{i=0}^{\infty} \alpha_i \eta(t-i)$$

and $\sum_{i=0}^{\infty} |\alpha_i|$ is convergent. There is no loss in generality in supposing that

$$\sum_{i=0}^{\infty} |\alpha_i| < 1.$$

Clearly
$$E(T_n) = \sum_{s=1}^{n} \sum_{i=0}^{\infty} \alpha_i E[\eta(t-s-i)]$$
$$= 0.$$

Write
$$\sigma^2 = E(\eta^2), \quad c_0 = E[X(t)^2], \quad c_s = E[X(t)X(t-s)].$$

Then
$$c_0 = \sigma^2(\alpha_0^2 + \alpha_1^2 + \ldots), \quad c_s = \sigma^2(\alpha_0\alpha_s + \alpha_1\alpha_{s+1} + \ldots),$$

which are both clearly convergent. Moreover,

$$R_n = E(T_n^2) = nE[X(t)^2] + 2\sum_{i=1}^{n-1}\sum_{s=1}^{n-i} E[X(t-i)X(t-i-s)]$$
$$= \left(nc_0 + 2\sum_{i=1}^{n-1}(n-i)c_i\right).$$

$n^{-1}R_n$ tends, as n increases, to
$$R_0 = \left(c_0 + 2\sum_{i=1}^{\infty} c_i\right)$$

if this series converges absolutely. We shall show that $\lim n^{-1}R_n$ is finite. For R_0 is clearly

$$\left(\sum_{i=0}^{\infty} \alpha_i\right)^2 \sigma^2, \tag{5}$$

and this is finite. Moreover, we notice that $n^{-1}R_n$ is not greater than

$$\left(\sum_{i=0}^{\infty} |\alpha_i|\right)^2 \sigma^2.$$

We must now impose the condition that $\sum_{i=0}^{\infty} \alpha_i$ is not zero. This condition is necessary to our

method of argument. If it is not zero, it may be assumed, without loss of generality, greater than a positive number. We now show that as n increases

$$pr\{t_0(2R_n)^{\frac{1}{2}} \leqslant T_n < t_1(2R_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $$\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\, dt.$$

We require the following lemma (Bernstein, 1927, p. 12):

LEMMA I. Let $$\rho_n = \Sigma_n + \sigma_n,$$
where ρ_n, Σ_n and σ_n are random variables such that

$$E(\Sigma_n) = E(\sigma_n) = 0, \quad E(\Sigma_n^2) = H_n, \quad E(\sigma_n^2) = H_n'.$$

Then if, for n large, $$pr\{t_0(2H_n)^{\frac{1}{2}} \leqslant \Sigma_n < t_1(2H_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $$\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\, dt,$$

then $$pr\{t_0(2J_n)^{\frac{1}{2}} \leqslant \rho_n < t_1(2J_n)^{\frac{1}{2}}\},$$

where $J_n = E(\rho_n^2)$ tends, uniformly in t_0 and t_1, to

$$\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\, dt,$$

provided that $$\lim_{n\to\infty} \frac{H_n'}{H_n} = 0.$$

Let ϵ be an arbitrarily small number and choose N so large that

$$\sum_{i=N}^{\infty} |\alpha_i| < \epsilon \sum_{i=0}^{\infty} |\alpha_i| < \epsilon.$$

Write $$X_1(t) = \sum_{i=0}^{N} \alpha_i \eta(t-i), \quad T_n' = \sum_{s=1}^{n} X_1(t-s).$$

Then $$E(T_n') = 0,$$

and write $$R_n' = E(T_n'^2).$$

We shall prove that the distribution of T_n' tends to normality, i.e. that

$$pr\{t_0(2R_n')^{\frac{1}{2}} \leqslant T_n' < t_1(2R_n')^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $$\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\, dt.$$

We first calculate R_n and R_n' in another way. For

$$\begin{aligned} T_n &= \sum_{s=1}^{n} X(t-s) = \sum_{s=1}^{n} \sum_{i=0}^{\infty} \alpha_i \eta(t-s-i) \\ &= \alpha_0 \eta(t-1) + (\alpha_0+\alpha_1)\eta(t-2) + \ldots + (\alpha_0 + \ldots + \alpha_{n-1})\eta(t-n) \\ &\quad + \sum_{s=1}^{\infty} (\alpha_s + \ldots + \alpha_{s+n-1})\eta(t-s-n), \end{aligned}$$

and so $R_n = E(T_n^2) = \sigma^2 \left\{\alpha_0^2 + (\alpha_0+\alpha_1)^2 + \ldots + (\alpha_0 + \ldots + \alpha_{n-1})^2 + \sum_{s=1}^{\infty} (\alpha_s + \ldots + \alpha_{s+n-1})^2\right\},$

and this series converges. On the other hand,

$$T'_n = \alpha_0\eta(t-1)+(\alpha_0+\alpha_1)\eta(t-2)+\ldots+(\alpha_0+\ldots+\alpha_{N-1})\eta(t-N)$$
$$+\sum_{p=1}^{n-N}(\alpha_0+\ldots+\alpha_N)\eta(t-N-p)$$
$$+(\alpha_1+\ldots+\alpha_N)\eta(t-n-1)+\ldots+\alpha_N\eta(t-n-N),$$

and so
$$R'^2_n = \sigma^2\{\alpha_0^2+\ldots(\alpha_0+\alpha_1)^2+\ldots+(\alpha_0+\ldots+\alpha_{N-1})^2$$
$$+(n-N)(\alpha_0+\ldots+\alpha_N)^2+(\alpha_1+\ldots+\alpha_N)^2+\ldots+\alpha_N^2\}.$$

Since we have already supposed that $\sum_{i=0}^{\infty}\alpha_i$ is positive, there exist positive numbers N_0 and d such that for all $N>N_0$, $\sum_{i=0}^{N}\alpha_i>d$. If this is not true the theorem is in general false. For suppose the distribution of the η's to be non-normal and write $\alpha_0=1$, $\alpha_1=-1$, $\alpha_i=0$ $(i>1)$. Then the distribution of T_n does not tend to normality and its variance does not increase with n. We shall later show that this condition on the α's is in fact satisfied for the solutions of stochastic difference equations.

Now by the ordinary central limit theorem, as n increases,

$$T''_n = \sum_{p=1}^{n-N}(\alpha_0+\ldots+\alpha_N)\eta(t-N-p)$$

tends to be distributed normally with zero mean and variance

$$R''_n = (n-N)(\alpha_0+\ldots+\alpha_N)^2\sigma^2,$$

that is
$$pr\{t_0(2R''_n)^{\frac{1}{2}}\leqslant T''_n<t_1(2R''_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to
$$\pi^{-\frac{1}{2}}\int_{t_0}^{t_1}e^{-t^2}\,dt.$$

Using Lemma I we see that the same is true, for fixed N, when we replace T''_n by T'_n and R''_n by R'_n. Now

$$T_n = T'_n+Q,$$

say, where Q is what we get if we replace the sequence $(\alpha_0,\alpha_1,\ldots)$ in T_n by $(\alpha_{N+1},\ldots)$ and alter t, and from (5) we can choose N so large that for $n>N$, $n^{-1}E(Q^2)<\epsilon$, say. Taking a sequence ϵ_1, ϵ_2, ... tending to zero and choosing first N sufficiently large and then n and using Lemma I again, we see that

$$pr\{t_0(2R_n)^{\frac{1}{2}}\leqslant T_n<t_1(2R_n)^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to
$$\pi^{-\frac{1}{2}}\int_{t_0}^{t_1}e^{-t^2}\,dt.$$

To complete the discussion we must show that the condition we have imposed on the sequence $\alpha_0,\alpha_1,\ldots$ is satisfied by the coefficients of the solutions of stationary stochastic difference equations. Consider an equation

$$X(t)+a_1X(t-1)+\ldots+a_hX(t-h) = \eta(t)$$

such that the roots of
$$z^h+a_1z^{h-1}+\ldots+a_h = 0 \qquad (6)$$

all lie inside the circle $|z|=1$. Then the solution of this equation is given (Wold, 1938, p. 53) by

$$X(t) = \sum_{i=0}^{\infty}\alpha_i\eta(t-i),$$

where the α_i are now the solutions of the infinite set of equations

$$\begin{array}{rl}
\alpha_0 = 1 & \\
a_1\alpha_0+\alpha_1 & = 0, \\
a_2\alpha_0+a_1\alpha_1 \quad +\alpha_2 & = 0, \\
\dots\dots\dots\dots\dots\dots\dots\dots & \\
a_h\alpha_0+a_{h-1}\alpha_1+\dots+\alpha_h & = 0, \\
a_h\alpha_1 \quad +\dots+a_1\alpha_h+\alpha_{h+1} & = 0, \\
\dots\dots\dots\dots\dots\dots\dots\dots &
\end{array}$$

and since the left-hand side is an absolutely convergent double series, we add, obtaining

$$(1+a_1+\dots+a_h)\sum_{i=0}^{\infty}\alpha_i = 1,$$

and so $\sum_{i=0}^{\infty}\alpha_i \neq 0$ and, as already observed, without loss of generality, may be supposed positive. This quantity is finite because all the roots of equation (6) lie inside the circle $|z| = 1$. Moreover, it follows that

$$R_0 = \left(\sum_{i=0}^{\infty}\alpha_i\right)^2\sigma^2 = (1+a_1+\dots+a_n)^{-2}\sigma^2.$$

This is, *in fact, proportional to the derivative at zero of the integrated power spectrum* (Wold, 1938, p. 69).

We now turn to the problem of discussing the relation between two such series and we consider the asymptotic distribution of

$$S_n = \sum_{t=1}^{n} X(-t)\,Y(-t),$$

where
$$X(t) = \sum_{i=1}^{\infty}\alpha_i\eta(t-i) \quad (\alpha_1 \neq 0), \tag{7}$$

and
$$Y(t) = \sum_{i=1}^{\infty}\beta_i\zeta(t-i) \quad (\beta_1 \neq 0). \tag{8}$$

We write S_n in this form rather than that of (3) for the sake of convenience in what follows, and we have altered the notation of the sums (7) and (8) so that they begin with the coefficients α_1 and β_1 for the same reason. Writing

$$c_s = E[X(t)\,X(t-s)], \quad d_s = E[Y(t)\,Y(t-s)] \quad (s = 0, 1, \dots),$$

as before, we have

$$c_s = \sigma_1^2(\alpha_1\alpha_{s+1}+\alpha_2\alpha_{s+2}+\dots), \quad d_s = \sigma_2^2(\beta_1\beta_{s+1}+\beta_2\beta_{s+2}+\dots),$$

where σ_1^2 and σ_2^2 are the second moments of η and ζ. Then

$$E(S_n) = \sum_{t=1}^{n} E(X(-t)\,Y(-t)) = 0,$$

$$\begin{aligned}
E(S_n^2) &= E\left\{\sum_{t=1}^{n} X(-t)\,Y(-t)\right\}^2 \\
&= E\left\{\sum_{t=1}^{n} X^2(-t)\,Y^2(-t)+2\sum_{t=1}^{n-1}\sum_{s=1}^{n-t} X(-t)\,X(-t-s)\,Y(-t)\,Y(-t-s)\right\} \\
&= nc_0d_0+2\sum_{s=1}^{n-1}(n-s)\,c_sd_s,
\end{aligned} \tag{9}$$

Consider the behaviour of $n^{-1}E(S_n^2)$ as n increases. Clearly

$$n^{-1}E(S_n^2) \to c_0 d_0 + 2\sum_{s=1}^{\infty} c_s d_s = C, \quad \text{say,} \tag{10}$$

if the series C is absolutely convergent. If X and Y are moving averages or the solutions of stationary stochastic difference equations this is certainly true, for in the first case the series is finite, and in the second it is majorized by a convergent geometric series. We show that it is true in the general case by the following argument. Without restricting generality, we may assume, as before, that $\sum_1^{\infty} |\alpha_i| < 1$, $\sum_1^{\infty} |\beta_i| < 1$. Then

$$\begin{aligned} |c_s| &\leqslant \sigma_1^2(|\alpha_1\alpha_s| + \dots) \\ &\leqslant \sigma_1^2(|\alpha_1| + \dots), \end{aligned}$$

and so

$$\begin{aligned} \left|\sum_1^{\infty} c_s d_s\right| &\leqslant \sum_1^{\infty} |c_s d_s| \leqslant \sigma_1^2 \sum_1^{\infty} |d_s| \\ &\leqslant \sigma_1^2\sigma_2^2 \sum_{s=1}^{\infty} (|\beta_1||\beta_s| + \dots) \\ &\leqslant \sigma_1^2\sigma_2^2 \left(\sum_{i=1}^{\infty} |\beta_i|\right)^2. \end{aligned} \tag{11}$$

Also

$$c_0 d_0 \leqslant \sigma_1^2\sigma_2^2 \left(\sum_{i=1}^{\infty} |\alpha_i|^2\right)\left(\sum_{i=1}^{\infty} |\beta_i|^2\right),$$

and so $c_0 d_0 + 2\sum_1^{\infty} c_s d_s$ is finite. We now prove that C is not zero. For

$$C = c_0 d_0 + 2\sum_{s=1}^{\infty} c_s d_s = \sigma_1^2\sigma_2^2\left[(\alpha_1^2 + \alpha_2^2 + \dots)(\beta_1^2 + \beta_2^2 + \dots) + 2\sum_{s=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \alpha_m\alpha_{m+s}\beta_n\beta_{n+s}\right],$$

and after some rearrangement, this equals

$$\sigma_1^2\sigma_2^2[(\alpha_1\beta_1)^2 + (\alpha_1\beta_2 + \alpha_2\beta_1)^2 + (\alpha_1\beta_3 + \alpha_2\beta_2 + \alpha_3\beta_1)^2 + \dots],$$

and $(\alpha_1\beta_1)^2$ is greater than zero and the rest non-negative at least. We therefore conclude that

$$n^{-1}E(S_n^2) \to C,$$

where

$$0 < C < \infty.$$

Assuming as before that $\sum_1^{\infty} |\alpha_i| < 1$, $\sum_1^{\infty} |\beta_i| < 1$,

we define N so that

$$\sum_{N+1}^{\infty} |\alpha_i| < \epsilon \sum_1^{\infty} |\alpha_i| < \epsilon, \tag{12}$$

$$\sum_{N+1}^{\infty} |\beta_i| < \epsilon \sum_1^{\infty} |\beta_i| < \epsilon, \quad \text{where } \epsilon \text{ is small.} \tag{13}$$

We now write

$$X_1(t) = \sum_{i=1}^{N} \alpha_i \eta(t-i), \tag{14}$$

$$Y_1(t) = \sum_{i=1}^{N} \beta_i \zeta(t-i), \tag{15}$$

and consider the sum

$$S_n' = \sum_1^{n} X_1(-t) Y_1(-t). \tag{16}$$

We begin by proving that when n is large this sum tends to be distributed in the normal form with a variance which is asymptotically equal to nC_1, where C_1 is obtained from C by putting $\alpha_i = \beta_i = 0$ for $i > N$. For it is then clearly true that

$$n^{-1}E(S_n'^2) \to C_1.$$

Now consider
$$S_n' = \sum_1^n X_1(-t)Y_1(-t),$$

where n is greater than N. For convenience of notation, we write

$$\eta_i = \eta(-i), \quad \zeta_i = \zeta(-i).$$

We then have
$$S_n' = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} A_{ij}\eta_i\zeta_j,$$

where the A_{ij} are certain constants. Moreover

$$\begin{aligned}
&E(\eta_i\zeta_j) = 0 && \text{all } i, j,\\
&E(\eta_i^2\zeta_j^2) = \sigma_1^2\sigma_2^2 && \text{all } i, j,\\
&E(\eta_i^2\zeta_j\zeta_k) = E(\eta_j\eta_k\zeta_i^2) = 0 && \text{for } j \neq k,\\
&E(\eta_i\eta_j\zeta_k\zeta_l) = 0 && \text{if } i \neq j \text{ or } k \neq l.
\end{aligned}$$

It therefore follows that
$$E(S_n'^2) = \sigma_1^2\sigma_2^2 \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} A_{ij}^2.$$

Inserting (14) and (15) in (16) we have

$$A_{ij} = 0$$

if
$$i > n+N, \quad \text{or} \quad j > n+N \quad \text{or} \quad |i-j| > N-1,$$

and
$$S_n' = \Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4,$$

where
$$\Sigma_1 = \sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij}\eta_i\zeta_j$$

with
$$\begin{aligned}
A_{ij} &= \alpha_i\beta_j + \alpha_{i-1}\beta_{j-1} + \dots + \alpha_{1+i-j}\beta_1 && \text{for } i > j,\\
&= \alpha_i\beta_j + \qquad\qquad + \alpha_1\beta_{1+j-i} && \text{for } i < j,\\
&= \alpha_1\beta_1 + \qquad\qquad + \alpha_i\beta_j && \text{for } i = j.
\end{aligned}$$

We also have
$$\Sigma_2 = \Sigma\Sigma A_{ij}\eta_i\zeta_j,$$

where the sum is taken over values of i and j such that $|i-j| < N$, $i \leq n$, $j \leq n$ and either $N < i$ or $N < j$, where

$$\begin{aligned}
A_{ij} &= \alpha_1\beta_{p+1} + \dots + \alpha_{N-p}\beta_N && \text{for } j-i = p > 0\\
&= \alpha_{p+1}\beta_1 + \dots + \alpha_N\beta_{N-p} && \text{for } i-j = p > 0\\
&= \alpha_1\beta_1 \quad + \dots + \alpha_N\beta_N && \text{for } i = j.
\end{aligned}$$

Then
$$E(\Sigma_2) = 0, \quad E(\Sigma_2^2) = \sigma_1^2\sigma_2^2\Sigma\Sigma A_{ij}^2,$$

where the sum is taken over the above values of i and j. This equals

$$(n-N)\sigma_1^2\sigma_2^2[(\alpha_1\beta_N)^2 + (\alpha_1\beta_{N-1} + \alpha_2\beta_N)^2 + \dots \\ + (\alpha_1\beta_1 + \dots + \alpha_N\beta_N)^2 + \dots + (\alpha_N\beta_2 + \alpha_{N-1}\beta_1)^2 + (\alpha_N\beta_1)^2]. \tag{17}$$

We know that $\alpha_1 \neq 0$. Let β_t be the first term of the sequence $\beta_N, \beta_{N-1}, \ldots, \beta_1$ which is not zero. Such a term certainly exists. Then the sum in the outer brackets of (17) will contain a term of the form $(\alpha_1\beta_t)^2$ and consequently $E(\Sigma_2^2) > 0$, and for N fixed will increase as $(n-N)$.

Next we have
$$\Sigma_3 = \Sigma\Sigma A_{ij}\eta_i\zeta_j,$$
where either
$$i \leqslant n, j > n \quad\text{and}\quad j-i<N, \tag{18}$$
or
$$j \leqslant n, i > n \quad\text{and}\quad i-j<N, \tag{19}$$
and
$$\begin{aligned} A_{ij} &= \alpha_{p+1}\beta_1 + \ldots + \alpha_N\beta_{N-p} \quad \text{for} \quad i-j=p>0 \\ &= \alpha_1\beta_{p+1} + \ldots + \alpha_{N-p}\beta_N \quad \text{for} \quad j-i=p>0. \end{aligned}$$
Then
$$E(\Sigma_3) = 0, \quad E(\Sigma_3^2) = \sigma_1^2\sigma_2^2\Sigma\Sigma A_{ij}^2,$$
where the sum is taken over the values (18) and (19).

Finally
$$\Sigma_4 = \sum_{i=n+1}^{n+N}\sum_{j=n+1}^{n+N} A_{ij}\eta_i\zeta_j,$$
where
$$\begin{aligned} A_{ij} &= \alpha_{i-n}\beta_{i-p-n} + \ldots + \alpha_N\beta_{N-p} \quad \text{for} \quad i-j=p>0 \\ &= \alpha_{j-p-n}\beta_{j-n} + \ldots + \alpha_{N-p}\beta_N \quad \text{for} \quad j-i=p>0 \\ &= \alpha_p\beta_p \qquad + \ldots + \alpha_N\beta_N \qquad \text{for} \quad i=j=n+p>n, \end{aligned}$$
and
$$E(\Sigma_4) = 0, \quad E(\Sigma_4^2) = \sigma_1^2\sigma_2^2\sum_{i=n+1}^{n+N}\sum_{j=n+1}^{n+N} A_{ij}^2.$$
We readily see that
$$E(\Sigma_i\Sigma_j) = 0 \quad \text{for} \quad i \neq j$$
and therefore
$$E(S_n'^2) = E(\Sigma_1^2) + E(\Sigma_2^2) + E(\Sigma_3^2) + E(\Sigma_4^2).$$
Moreover, for constant N, $E(\Sigma_1^2)$, $E(\Sigma_3^2)$ and $E(\Sigma_4^2)$ are constant, and so for large n we have
$$\begin{aligned} n^{-1}E(S_n'^2) \to C_2 = \sigma_1^2\sigma_2^2[(\alpha_1\beta_N)^2 + (\alpha_1\beta_{N-1} + \alpha_2\beta_N)^2 + \ldots \\ + (\alpha_1\beta_1 + \ldots + \alpha_N\beta_N)^2 + \ldots + (\alpha_N\beta_1)^2] \neq 0. \end{aligned} \tag{20}$$

Now suppose that N is fixed and consider the sum $\sum_1^n X_1(-t)Y_1(-t)$. We write
$$n = m(m+N) + p,$$
where $p < 2m+N+1$ and n is large enough for m to be greater than N. This equation fixes m which increases roughly as $n^{\frac{1}{2}}$ when n increases. Write
$$\begin{aligned} S_n' &= \left(\sum_{t=1}^{N} + \sum_{N+1}^{N+m} + \ldots + \sum_{1+m(N+m-1)}^{m^2+mN} + \sum_{n-p+1}^{n}\right) X_1(-t)Y_1(-t) \\ &= V_1 + U_1 + V_2 + U_2 + \ldots + V_m + U_m + W. \end{aligned}$$
Then $V_1, \ldots, V_m$ and W are all independent and $E(V_1^2), \ldots, E(V_m^2)$ are independent of n, and in fact not greater than KN, where K is a constant independent of N. Also $E(W^2)$ is not greater than $K(2m+N+1)$, where K may be taken as the same constant. $U_1, \ldots, U_m$ are also all independent and $E(U_m^2)$ is asymptotically equal to mC_2 when n (and therefore m) are large. Therefore, writing
$$A_m = U_1 + \ldots + U_m, \quad B_m = V_1 + \ldots + V_m + W,$$

we have
$$E(A_m) = 0, \quad E(A_m^2) = \sum_{i=1}^{m} E(U_i^2),$$
$$E(B_m) = 0, \quad E(B_m^2) = \sum_{i=1}^{m} E(V_i^2) + E(W^2),$$
and the latter increases as m, whilst the former increases as m^2 and so
$$E(B_m^2)\{E(A_m^2)\}^{-1} \to 0$$
as n increases.

By Lemma I it is therefore sufficient to show that the distribution of A_m tends to normality.

LEMMA II (Liapounoff's Central Limit Theorem, Bernstein, 1927). If
$$\Sigma_m = u_1^{(m)} + \dots + u_m^{(m)}$$
is the sum of m independent quantities such that
$$E(u_r^{(m)}) = 0, \quad E(u_r^{(m)2}) = b_r^{(m)}, \quad E(u_r^{(m)4}) = c_r^{(m)},$$
and if, as m increases,
$$b_m^{-2} \sum_{r=1}^{m} c_r^{(m)} \to 0,$$
where
$$b_m = \sum_{r=1}^{m} b_r^{(m)} = E(\Sigma_m^2),$$
then
$$pr\{t_0(2b_m)^{\frac{1}{2}} \leqslant \Sigma_m < t_1(2b_m)^{\frac{1}{2}}\}$$
tends, uniformly in t_0 and t_1, to
$$\pi^{-\frac{1}{2}} \int_{t_0}^{t_1} e^{-t^2}\, dt.$$

To apply the lemma we put $U_r = u_r^{(m)}$. We already have $E(U_r) = 0$. Also
$$m^{-1}E(U_r^2) \to C_2 > 0 \quad \text{by (20)},$$
and so
$$m^{-2}b_m \to C_2.$$
Now consider
$$c_r^{(m)} = E(U_r^4),$$
where
$$U_r = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} A_{pq}\, \eta_{r(N+r-1)+p-1}\, \zeta_{r(N+r-1)+q-1},$$
and the A_{pq} are calculated with m in place of n. Since the η's all have the same probability distribution and similarly for the ζ's, we shall write η_p and ζ_q for $\eta_{r(N+r-1)+p-1}$ and $\zeta_{r(N+r-1)+q-1}$ for the sake of convenience. So we can write the above
$$U_r = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} A_{pq}\, \eta_p\, \zeta_q.$$

U_r^4 will be a polynomial of the fourth order in the η's and the ζ's and its expectation may be regarded as the sum of two distinct types of terms so that $E(U_r^4) = \Sigma E(w_1) + \Sigma E(w_2)$, where the terms w_1 are of the form $A_{pq}^4 \eta_p^4 \zeta_q^4$, and the terms w_2 are of the form $A_{pq}^2 A_{kl}^2 \eta_p^2 \zeta_q^2 \eta_k^2 \zeta_l^2$ with $(p, q) \neq (k, l)$. All other terms arising in the product will clearly vanish when the expectation is taken.

Then, since the A_{pq} are bounded and the number of non-zero terms in w_1 and w_2 are not greater than $2N(m+N)$ and $4N^2(m+N)^2$ respectively, we have
$$E(U_r^4) < Km^2,$$
where K is a constant depending on N but independent of m and n. It follows that
$$b_m^{-2} \sum_{1}^{m} C_r^{(m)}$$

is of order m^{-1} and tends to zero as n and m increase. The conditions of the lemma are therefore satisfied and we conclude that

$$pr\{t_0(2E(A_m^2))^{\frac{1}{2}} \leqslant A_m < t_1(2E(A_m^2))^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to $\qquad \pi^{-\frac{1}{2}}\int_{t_0}^{t_1} e^{-t^2}\,dt.$

Applying Lemma I we have

$$pr\{t_0[2E(S_n'^2)]^{\frac{1}{2}} \leqslant S_n' < t_1[2E(S_n'^2)]^{\frac{1}{2}}\}$$

tends, uniformly in t_0 and t_1, to the same limit.

We now consider the relationship between S_n' and S_n. Write

$$\begin{aligned}
S_n'' = S_n - S_n' &= \sum_1^n X(-t)\,Y(-t) - \sum_1^n X_1(-t)\,Y_1(-t) \\
&= \sum_{t=1}^n \left(\sum_{i=1}^\infty \alpha_i \eta_{t+i}\right)\left(\sum_{j=1}^\infty \beta_j \zeta_{t+j}\right) - \sum_{t=1}^n \left(\sum_{i=1}^N \alpha_i \eta_{t+i}\right)\left(\sum_{j=1}^N \beta_j \zeta_{t+j}\right) \\
&= \sum_{t=1}^n \left(\sum_{i=N+1}^\infty \alpha_i \eta_{t+i}\right)\left(\sum_{j=N+1}^\infty \beta_j \zeta_{t+j}\right) \\
&\quad + \sum_{t=1}^n \left(\sum_{i=N+1}^\infty \alpha_i \eta_{t+i}\right)\left(\sum_{j=1}^N \beta_j \zeta_{t+j}\right) \\
&\quad + \sum_{t=1}^n \left(\sum_{i=1}^N \alpha_i \eta_{t+i}\right)\left(\sum_{j=N+1}^\infty \beta_j \zeta_{t+j}\right) \\
&= W_1 + W_2 + W_3. \qquad (21)
\end{aligned}$$

We must now calculate the variance of these terms. Consider again (9). We have shown (11) that

$$\begin{aligned}
2\sum_1^\infty c_s d_s &\leqslant 2\sigma_1^2\sigma_2^2(|\beta_0| + |\beta_1| + \ldots)^2 \\
&\leqslant 2\sigma_1^2\sigma_2^2(|\alpha_0| + |\alpha_1| + \ldots)^2,
\end{aligned}$$

and we now apply this to the three sums in equation (21). It follows that if N be chosen to satisfy the conditions (12) and (13) then

$$\overline{\lim}\, n^{-1}E(W_1^2) < K\sigma_1^2\sigma_2^2\epsilon^2, \quad \overline{\lim}\, n^{-1}E(W_2^2) < K\sigma_1^2\sigma_2^2\epsilon^2, \quad \overline{\lim}\, n^{-1}E(W_3^2) < K\sigma_1^2\sigma_2^2\epsilon^2,$$

where K is a constant independent of N.

It follows that

$$S_n = S_n' + W_1 + W_2 + W_3,$$

where the variance of W_1, W_2 and W_3 can be made small compared with that of S_n by choosing N large. Then by first choosing N large and then n and using Tchebycheff's inequality, we see that the distribution of S_n tends to normality with variance $E(S_n^2)$ and this completes the proof.

In the general application of the above results some care is needed. We can suppose that our empirical values of X and Y are distributed about their sample means which we take to be zero and we must estimate the variance of S_n from formulae (9), or (approximately) from (10). But we must not insert in this formula the sample covariances for the c_s and the d_s because, as Bartlett (1946) has shown, the standard errors of the sample values of these covariances are of order $n^{-\frac{1}{2}}$ and we cannot therefore expect the series (10) to converge, let

alone give the correct value. To use the formula correctly we must first decide on the order and coefficients of the stochastic difference equation which we can suppose generated the series and, from these coefficients, calculate the value of (9).

In the case where the series are generated by a three-term difference equation, the calculations are simplified. Suppose the X and Y satisfy the equations

$$X(t+2)+aX(t+1)+bX(t)=\eta(t+2),$$
$$Y(t+2)+AY(t+1)+BY(t)=\zeta(t+2),$$

where
$$E(\eta(t))=E(\zeta(t))=0$$

and
$$E(\eta^2(t))=\sigma_1^2,\quad E(\zeta^2(t))=\sigma_2^2,$$

as before. For the series to be stationary, we must have $b<1$, $B<1$. We suppose that in addition to this the series are oscillatory and so $a^2<4b$, $A^2<4B$. The solutions will then be

$$X(t)=\sum_{s=0}^{\infty}2(4b-a^2)^{-\frac{1}{2}}p^s\sin\theta s\,\eta(t-s+1),$$
$$Y(t)=\sum_{s=0}^{\infty}2(4B-A^2)^{-\frac{1}{2}}P^s\sin\phi s\,\zeta(t-s+1),$$

where $p=b^{\frac{1}{2}}$, $P=B^{\frac{1}{2}}$, $\cos\theta=-a(2b^{\frac{1}{2}})^{-1}$, $\cos\phi=-A(2B^{\frac{1}{2}})^{-1}$. Also (Kendall, 1946, p. 408)

$$r_s=\frac{c_s}{c_0}=\frac{p^s\sin(s\theta+\psi)}{\sin\psi},\quad R_s=\frac{d_s}{d_0}=\frac{P^s\sin(s\phi+\Phi)}{\sin\Phi},$$

where
$$\tan\psi=\frac{1-p^2}{1+p^2}\tan\theta\quad\text{and}\quad\tan\Phi=\frac{1-P^2}{1+P^2}\tan\phi$$

and
$$c_0=\sigma_1^2\frac{1+b}{(1-b)\{(1+b)^2-a^2\}},\quad d_0=\sigma_2^2\frac{1+B}{(1-B)\{(1+B)^2-A^2\}}.$$

We then need to calculate

$$\begin{aligned}C&=c_0d_0+2\sum_{s=1}^{\infty}c_sd_s\\&=c_0d_0\left\{1+2\sum_{s=1}^{\infty}r_sR_s\right\}\\&=c_0d_0\left\{1+\sum_{s=1}^{\infty}\frac{p^sP^s\sin(s\theta+\psi)\sin(s\phi+\Phi)}{\sin\psi\,\sin\Phi}\right\}\\&=c_0d_0\left\{1+\frac{2pP}{\sin\psi\,\sin\Phi}\left[\frac{\cos(\psi-\Phi+\theta-\phi)-pP\cos(\psi-\Phi)}{1-2pP\cos(\theta-\phi)+p^2P^2}\right.\right.\\&\qquad\left.\left.-\frac{\cos(\psi+\Phi+\theta+\phi)-pP\cos(\psi+\Phi)}{1-2pP\cos(\theta+\phi)+p^2P^2}\right]\right\}.\qquad(22)\end{aligned}$$

It is probably easiest to calculate C from this equation rather than attempt to simplify (22) still further. I hope to discuss the practical application of these formulae in another paper.

REFERENCES

WOLD, H. (1938). *A Study in the Analysis of Stationary Time Series*. Uppsala.
BERNSTEIN, S. (1927). *Math. Ann.* **97**, 1.
BARTLETT, M. S. (1946). *Supp. J. Roy. Statist. Soc.* **8**, 27.
KENDALL, M. G. (1946). *Advanced Theory of Statistics*, 2. London: Charles Griffin and Co.

RANK CORRELATION BETWEEN TWO VARIABLES, ONE OF WHICH IS RANKED, THE OTHER DICHOTOMOUS

By J. W. WHITFIELD, *Psychological Laboratory, University of Cambridge*

Rank correlation is one of the most useful statistical techniques available for the treatment of data arising in experimental and applied psychological research. Chambers (1946) has indicated the type of data most frequently occurring in these fields, and has pointed out the advantages of Kendall's τ over Spearman's ρ or any form of transformation to ordinal form.

Given the use of τ when tied rankings are present (Kendall, 1946) it seemed possible to extend the method to cover a very common problem in psychology, namely, determination of the relation between two variables, one of which is expressed as a ranking and the other as a dichotomy. In applied or field work the relation of a psychological 'measurement' and an external criterion nearly always appears in this form. The usual method of determining the relationship consists of reducing the ranking to a dichotomy and calculating χ^2 for the 2×2 table which results. That this may lead to inaccuracy can be seen from the following example:

Variable A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Variable B	+	+	+	+	+	+	+	−	−	−	+	+	+	−	−	−	−	−	−	−
Variable C	−	−	−	+	+	+	+	+	+	+	−	−	−	−	−	−	−	+	+	+

Here the data are supposed to be ranked according to variable A and dichotomized into + and − with respect to variables B and C.

Treating the relation between variables A and B as a 2×2 contingency table:

	Variable B	
	+	−
Variable A: Rankings 1–10	7	3
Rankings 11–20	3	7

Applying χ^2, P is found to be 0·074 without Yates's correction for continuity, or 0·180 if the correction is applied.

But χ^2 is exactly the same for the contingency table relating variables A and C, although it is obvious from the data that there is considerable difference in the two relationships, the evidence for which is sacrificed by reducing the ranking to a dichotomy.

If, alternatively, we consider the dichotomous variable as a ranking composed entirely of two sets of tied rankings, we may calculate the coefficients between A and B, A and C respectively which I shall denote by τ_{AB}, τ_{AC}. The corresponding values of S will be found to be, after the manner described by Kendall (1946):

$$S = +70 - 9 + 21 = +82,$$

$$S = -30 + 49 - 21 = -2.$$

For the calculation of τ in the case of tied rankings we have a choice in the denominator by which S is to be divided to give τ. In the untied case this would be $\frac{1}{2}n(n-1)$, where n is

the number of ranks. In the tied case we may take the denominator S' as $\frac{1}{2}n(n-1)$ or as $[\frac{1}{2}n(n-1)\{\frac{1}{2}n(n-1)-\frac{1}{2}\Sigma t(t-1)\}]^{\frac{1}{2}}$, where t_1, t_2, etc., are the extent of the ties. The choice is determined by practical considerations (see Kendall, 1946), but is not material to a discussion of significance. For an untied ranking and a dichotomy with 'x' and 'y' members in each class, the second form reduces to $\{\frac{1}{2}xyn(n-1)\}^{\frac{1}{2}}$.

In the case of two untied rankings Kendall has shown that $\operatorname{var} S = \frac{1}{18}n(n-1)(2n+5)$. In the case of one untied ranking and one with ties of extent t_1, t_2, etc., Sillitto (1947) has extended this result by proving that

$$\operatorname{var} S = \tfrac{1}{18}\{n(n-1)(2n+5)-\Sigma t(t-1)(2t+5)\}. \tag{1}$$

In the case of an untied ranking and a dichotomy, $t_1 = x$, $t_2 = y$, $(x+y) = n$, and we have then the simple form

$$\operatorname{var} S = \tfrac{1}{3}xy(n+1). \tag{2}$$

In the example above this gives

$$\operatorname{var} S = \frac{(10)(10)(21)}{3} = 700, \quad \sqrt{(\operatorname{var} S)} = 26\cdot46, \quad \frac{S_{AB}}{\sqrt{(\operatorname{var} S)}} = \frac{82-1}{26\cdot46} = 3\cdot06.$$

The probability of a deviation greater than this in absolute value is 0·0022. Further,

$$\frac{S_{AC}}{\sqrt{(\operatorname{var} S)}} = \frac{2-1}{26\cdot46} = 0\cdot0378,$$

and the corresponding probability is 0·970.

VARIANCE WHEN THERE ARE TIES IN THE RANKING

The variance of S given by equation (2) is true only in the case of a dichotomy and an untied ranking. For a tied ranking I surmised from some special cases that

$$\operatorname{var} S = \frac{xy}{3n(n-1)}\{(n^3-n)-\Sigma(t^3-t)\}. \tag{3}$$

In the note following this paper Mr Kendall provides proof of this result.

Example (from data collected by the Medical Research Council team in Germany 1946, as yet unpublished). Selected workers in a factory were interviewed and an assessment made of their adaptation to living conditions. They were assessed as 'Efficient' or 'Overactive'. Other data were available, including statements by the men of frequency of nocturia. For men aged 50–59 years the following was observed:

Assessment	*Rank order of frequency of nocturia* (least frequent nocturia given highest rank)
Efficient	$2\frac{1}{2}$, $2\frac{1}{2}$, $2\frac{1}{2}$, $2\frac{1}{2}$, $6\frac{1}{2}$, $6\frac{1}{2}$, 10, 10, 10, 10, 14, 14
Overactive	5, 10, 14, 16, 17

Five is the highest ranking in the overactive group. Four members of the efficient group have higher rankings, and eight lower rankings. The S score for that member is therefore $4-8$. Similarly, for all members we have

$$S = 4-8+6-2+10+12+12 = +34.$$

Using a denominator in the form

$$[xy\{\tfrac{1}{2}n(n-1)-\tfrac{1}{2}\Sigma t(t-1)\}]^{\frac{1}{2}},$$

τ is given by

$$\tau = +\frac{34}{\sqrt{[(12)(5)\{\frac{1}{2}(17)(16)-\frac{1}{2}(4)(3)-\frac{1}{2}(2)(1)-\frac{1}{2}(5)(4)-\frac{1}{2}(3)(2)\}]}} = +\frac{34}{\sqrt{6960}} = +0{\cdot}408.$$

From (3) we then have

$$\begin{aligned}\operatorname{var} S &= \frac{(12)(5)}{3(17)(16)}\{(17^3-17)-(4^3-4)-(2^3-2)-(5^3-5)-(3^3-3)\}\\ &= 344{\cdot}6.\end{aligned}$$

A small problem arises when we consider the correction for continuity to be applied in testing the significance of an observed value of S. In the case of a dichotomy and an untied ranking the interval between successive S values is 2. In the case of a dichotomy and a ranking composed entirely of ties of the same extent 't', the interval is $2t$. But in the example the ties are of varying extent, and the interval between successive S values is composed of a mixture of the intervals produced by the successive rank values. Thus, although these varying intervals are combined so that over most of the range the interval between successive values is unity, the distribution oscillates somewhat, and to use the value $\frac{1}{2}$ as the correction for continuity would sometimes be misleading. Further work is required to determine the correction which will provide a probability on the normal distribution equal to or slightly greater than the true probability in all cases. Until this is available I propose to use a crude correction, based on the average of the intervals mentioned above. In the example the successive rank values $2\frac{1}{2}$ and 5 give an interval of 5 in S score, rank values 5 and $6\frac{1}{2}$ give an interval of 3, and it is therefore possible to determine the average interval by calculating the intervals given by successive rank values. This calculation can be shortened. The total of the S score intervals is twice the number of members, less the extent of the ties involving the first and last members. If we divide this by the number of intervals between successive rank values we have the average S score interval. In the example this is $\frac{1}{6}(34-4-1)$. Using half of this as the correction for continuity we have

$$\frac{S}{\sqrt{(\operatorname{var} S)}} = \frac{34-2{\cdot}42}{18{\cdot}56} = 1{\cdot}702.$$

The pre-observational hypothesis, made on psychological grounds, was that excessive nocturia is a symptom of inefficient adaptation to living conditions, i.e. a positive correlation should be obtained. From these observations the probability of a positive correlation as great or greater than the observed value appearing by chance is 0·044. Direct calculation of the positive tail of the distribution of S gives a probability of 0·0368.

The alternative testing hypothesis based on the absolute value of S gives a probability twice as great, and the corresponding direct calculation using both positive and negative tails of the actual S distribution gives a probability of 0·0735.

By itself this evidence could only be debatable substantiation of the psychological hypothesis. In fact, additional data from two other factory groups, treated in the same way, gave a total S value of $+104$, the square-root of the total variance being 35·00, providing a justification of the hypothesis.

THE CASE OF THE 2×2 TABLE

If one dichotomous variable can be considered as a ranking with two sets of tied ranks it is logical to consider the case when both variables are in this form. If we have a 2×2 table in the form

(AB)	(Ab)	(A)
(aB)	(ab)	(a)
(B)	(b)	N

any member of (AB) taken with any member of (ab) has the same order in either ranking and hence contributes $+1$ to S, and any member of (Ab) with any member of (aB) contributes -1. The others contribute nothing. Hence

$$S = (AB)(ab) - (Ab)(aB).$$

From equation (3)

$$\begin{aligned} \operatorname{var} S &= \frac{(A)(a)}{3N(N-1)}[(N^3 - N) - \{(B)^3 - (B)\} - \{(b)^3 - (b)\}] \\ &= \frac{(A)(a)(B)(b)}{N-1}. \end{aligned} \tag{4}$$

Again, for testing the significance of an observed value of S it is necessary to correct for continuity by subtracting half the interval between successive S values. In the case of the 2×2 table the interval is N, for if we increase (AB) by unity S becomes

$$\{(AB)+1\}\{(ab)+1\} - \{(Ab)-1\}\{(aB)-1\} = (AB)(ab) - (Ab)(aB) + N.$$

Hence, for the normal deviate, we have

$$\frac{S - \frac{1}{2}N}{\sqrt{\left\{\frac{(A)(a)(B)(b)}{N-1}\right\}}}. \tag{5}$$

It will be noted that τ (taking the ties into account in calculating the denominator S') is

$$\frac{(AB)(ab) - (Ab)(aB)}{\sqrt{[(A)(a)(B)(b)]}},$$

which is the product-moment correlation for a 2×2 table when the variables are conventionally regarded as possessing the discrete values 0, 1.

Testing by use of the normal deviate seems to be moderately accurate, and would appear to be useful in those cases where χ^2 is suspect because of small expectations in the cells of the 2×2 table. It is less laborious to calculate than the hypergeometric treatment, and is an alternative form of the approximation to hypergeometric treatment given by Pearson (1947), who also discusses the order of accuracy of the approximation.

Using the data given earlier as an example, but assuming that it had been possible only to grade nocturia into 'Normal' or 'Excessive', we have the following table:

Assessment	Nocturia	
	Normal	Excessive
Efficient	10	2
Overactive	2	3

$$S = 30 - 4 = 26, \quad \text{var}\, S = \frac{(12)(5)(12)(5)}{16} = 225.$$

This gives, after correction for continuity,

$$\frac{S - \frac{1}{2}N}{\sqrt{(\text{var}\, S)}} = \frac{26 - 8\cdot 5}{15} = 1\cdot 1667.$$

This gives the probability of S being attained or exceeded in the direction of the hypothesis (i.e. positive values only) as 0·1217. χ^2 without the continuity correction gives $P = 0\cdot 0369$,* and with the correction, $P = 0\cdot 1143$. The hypergeometric treatment, summing the probabilities of obtaining 3, 4 or 5 in the Overactive-Excessive category, gives $P = 0\cdot 1166$.

If the more customary test of absolute value is applied, χ^2 with Yates's correction gives $P = 0\cdot 2286$, S and the normal deviate gives $P = 0\cdot 2434$, i.e. both values of P are doubled. The hypergeometric treatment, adding the probability of obtaining 0 in the Overactive-Excessive category gives $P = 0\cdot 2445$.

It will be seen that in conditions such as these, S and the normal deviate give a reasonable approximation to the exact treatment.

* This is making the common assumption that $\{(AB)(ab) - (Ab)(aB)\}^2 N/\{(A)(a)(B)(b)\}$ is distributed as χ^2 with 1 degree of freedom, or that its square root is a normal deviate with sign depending on the sign of $(AB)(ab) - (Ab)(aB)$.

REFERENCES

Chambers, E. G. (1946). Statistical techniques in applied psychology. *Biometrika*, **33**, 269.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81.

Kendall, M. G. (1946). The treatment of ties in ranking problems. *Biometrika*, **33**, 239.

Pearson, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. *Biometrika*, **34**, 139.

Sillitto, G. P. (1947). The distribution of Kendall's coefficient of rank correlation in rankings containing ties. *Biometrika*, **34**, 36.

THE VARIANCE OF τ WHEN BOTH RANKINGS CONTAIN TIES

By M. G. KENDALL

1. The variance of τ in the population of sample permutations was given in my paper of 1938 for the case where no tied ranks exist. Mr Sillitto (1947) has given the formula where one ranking contains ties but the other does not. In the foregoing paper Mr Whitfield has correctly surmised the variance when one ranking contains ties, and the other is a dichotomy. In this note I derive the general formula for the variance when both rankings contain ties. The results of Messrs Sillitto and Whitfield then follow as special cases.

2. I shall follow the method of Daniels (1944). If a_{ij} represents the contribution of the ith and jth members of a ranking to τ we have

$$\begin{aligned} a_{ij} &= +1 \quad (i<j) \\ &= 0 \quad (i=j) \\ &= -1 \quad (i>j) \end{aligned} \tag{1}$$

We write
$$c_{ij} = a_{ij} b_{ij}, \tag{2}$$
where a and b refer to different rankings, and
$$c = \sum_{i,j=1}^{n} c_{ij}. \tag{3}$$
The quantity c is simply related to S by the relation
$$c = 2S, \tag{4}$$
and for the testing of τ it is sufficient to test c or S which are merely constant multiples of τ. I work with the quantity c.

3. We have, from Daniel's results,

$$\sum_{l=1}^{n} a_{il} = n+1-2i, \tag{5}$$

$$\sum_{i,l=1}^{n} a_{il}^2 = n(n-1), \tag{6}$$

$$\sum_{i,l,t=1}^{n} a_{il}\, a_{tl} = \tfrac{1}{3}n(n^2-1), \tag{7}$$

$$E(c) = 0, \tag{8}$$

$$E(c^2) = \frac{4}{n(n-1)(n-2)}\{\Sigma a_{il}a_{tl} - \Sigma a_{il}^2\}\{\Sigma b_{il}b_{tl} - \Sigma b_{il}^2\} + \frac{2}{n(n-1)}\Sigma a_{il}^2 \Sigma b_{il}^2. \tag{9}$$

If we substitute from (6) and (7) in (9) we find
$$E(c^2) = \tfrac{2}{9}n(n-1)(2n+5), \tag{10}$$
or, equivalently,
$$E(S^2) = \tfrac{1}{18}n(n-1)(2n+5), \tag{11}$$
from which the variance of τ in the case of untied rankings follows at once.

4. Now suppose that sets of t_1, t_2, ... consecutive members in one ranking are tied. In place of (6) we then have
$$\sum_{i,l=1}^{n} a_{il}^2 = n(n-1) - \Sigma\, t(t-1), \tag{12}$$

the summation on the right taking place over the various values of t. This result follows simply from the consideration that for a pair of tied ranks $a_{ij} = 0$, and consequently the sum of squares of contributions from a tied set is of the same form as for the ranking as a whole.

In place of (7) we have
$$\sum_{i,l,t=1}^{n} a_{il}a_{it} = \tfrac{1}{3}n(n^2-1) - \tfrac{1}{3}\Sigma\, t(t^2-1). \tag{13}$$

This is not quite so obvious. Consider a set of tied ranks. The contribution to the sum on the left of (13) will be unchanged if the suffixes l, t fall outside this set. If they both fall inside, no contribution arises and therefore we have to subtract the term $\frac{1}{3}t(t^2-1)$. The remaining possibility is that one falls inside and one outside. In such a case the contribution remains unchanged in total for it is zero in the original untied case, each possible pair occurring once to give $+1$ and one to give -1. Formula (13) follows.

5. By substitution in (9) we then have, for two rankings with ties typified respectively by t and u,
$$\begin{aligned} E(c^2) = {}& \frac{4}{n(n-1)(n-2)}\{\tfrac{1}{3}n(n-1)(n-2) - \tfrac{1}{3}\Sigma t(t-1)(t-2)\} \\ & \times\{\tfrac{1}{3}n(n-1)(n-2) - \tfrac{1}{3}\Sigma u(u-1)(u-2)\} \\ & + \frac{2}{n(n-1)}\{n(n-1) - \Sigma t(t-1)\}\{n(n-1) - \Sigma u(u-1)\}. \end{aligned} \tag{14}$$

This is the general formula required. We can express it in the alternative form
$$\begin{aligned} E(c^2) = {}& \tfrac{2}{9}n(n-1)(2n+5) - \tfrac{2}{9}\Sigma t(t-1)(2t+5) - \tfrac{2}{9}\Sigma u(u-1)(2u+5) \\ & + \frac{4}{9n(n-1)(n-2)}\{\Sigma t(t-1)(t-2)\}\{\Sigma u(u-1)(u-2)\} \\ & + \frac{2}{n(n-1)}\{\Sigma t(t-1)\}\{\Sigma u(u-1)\}. \end{aligned} \tag{15}$$

6. (i) If one ranking is untied, say all the u's are zero, we have Mr Sillitto's result
$$E(c^2) = \tfrac{2}{9}n(n-1)(2n+5) - \tfrac{2}{9}\Sigma t(t-1)(2t+5). \tag{16}$$

(ii) If one ranking is untied and the other is a dichotomy into x and $n-x=y$ members, (16) reduces to
$$E(c^2) = \tfrac{4}{3}xy(n+1), \tag{17}$$
agreeing with Mr Whitfield's equation (2).

(iii) If one ranking contains ties and the other is a dichotomy we find on substitution in (14)
$$E(c^2) = \frac{4xy}{3n(n-1)}\{n^3 - n - \Sigma(t^3-t)\}, \tag{18}$$
agreeing with Mr Whitfield's equation (3).

(iv) Finally, if both variates are dichotomized into x, y and p, q we find
$$E(c^2) = \frac{4xypq}{n-1}, \tag{19}$$
agreeing with Mr Whitfield's equation (4).

REFERENCES

See the references to Mr Whitfield's paper together with:

DANIELS, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, **33**, 129.

into the ith group, and let m_i $(i = 1, 2, \ldots, k)$ be the expected number. It is possible theoretically for χ^2 to be calculated for the case where

$$\sum_{i=1}^{k} n_i = N \neq \sum_{i=1}^{k} m_i,$$

but such cases must be rare in statistical practice. We shall overlook this case and will consider the case where the totals of observed and expected are made equal to one another with the resultant loss of one degree of freedom in the calculation of χ^2. If the totals agree then

$$\sum_{i=1}^{k} n_i - \sum_{i=1}^{k} m_i = \sum_{i=1}^{k} \delta_i = 0,$$

where $\delta_i = n_i - m_i$. In order that the sum of these δ's should be zero, at least one of them must be negative in sign, but which one of these δ's it will be would seem to be a matter of chance. It is on this fact that we shall base the first test criterion.

4. Suppose that we have a sequence of signs of which r_1 are positive and r_2 negative, where $r_1 + r_2 = r$, and $r_1 > 0$ and $r_2 > 0$. These signs are postulated to occur in a random order. Given such a sequence it is easy to record the number of sets of positive and negative signs. For example, if the sequence is

$$+\,+\,-\,-\,+\,+\,-\,-\,+\,-\,+\,+\,+\,+\,-,$$

then $r = 15$, $r_1 = 9$, $r_2 = 6$, and there are four sets of positive signs and four sets of negative signs. In general there can be (α) t positive, t negative, or (β) t positive, $t+1$ negative, or (γ) $t+1$ positive and t negative sets of signs. If $T = 2t$ or $2t+1$ as required, we may ask what is the probability that given r_1 and r_2 such a number T of sets (alternately positive and negative) would have arisen through chance. This probability follows at once from Whitworth, *Choice and Chance*, Proposition XXV, viz.: 'The number of ways in which n indifferent things can be distributed in t different parcels (blank lots being inadmissible) is

$$(n-1)!/(t-1)!\,(n-t)!\text{'.*}$$

5. The total number of ways in which r_1 and r_2 elements can be arranged is

$$\frac{r!}{r_1!\,r_2!}.$$

We now require to enumerate the number of ways in which r_1 can be arranged to form t sets and r_2 to form t sets. To arrange r_1 in t sets is equivalent (vide Whitworth) to making $t-1$ breaks in a sequence of r_1 observations, and this may be done in

$$(r_1-1)!/(t-1)!\,(r_1-t)! \text{ ways,}$$

and similarly for r_2. It is not specified whether + or − should start the sequence, and hence the total number of ways in which a sequence $r_1 + r_2$ may be arranged in t sets each is

$$\frac{2(r_1-1)!\,(r_2-1)!}{(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!}.$$

* Since I first thought of this method of attack I have found that the distribution of groups as given by me in § 5 has already been given by W. L. Stevens, *Ann. Eugen., Lond.*, **9**, 10, and by A. Wald and J. Wolfowitz, *Ann. Math. Statist.* **11**, 147. The probability function has been tabled by F. S. Swed and C. Eisenhart, *Ann. Math. Statist.* **14**, 66, but it is not in a form that I found suitable for my purposes. The probability function has been known for many years; what is interesting is the different uses to which it has been put.

The probability of $2t$ sets will be

$$P\{2t \mid r_1, r_2\} = \frac{2r_1!(r_1-1)!\,r_2!(r_2-1)!}{r!(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!}, \tag{1}$$

and the probability of obtaining $(2t+1)$ sets will be

$$P\{2t+1 \mid r_1, r_2\} = P\{t \mid r_1, t+1 \mid r_2\} + P\{t+1 \mid r_1, t \mid r_2\} = P\{2t \mid r_1, r_2\}\left(\frac{r-2t}{2t}\right). \tag{2}$$

Hence given r_1, r_2 and T from a random sequence of positive and negative signs the probability of such a number of sets having arisen through chance may be calculated.

6. It is desired to use the probability of a given arrangement of signs in order to test a given hypothesis represented by a *smooth* probability law, bearing in mind that, if the given hypothesis is not true, then any alternative law is likely to be of a smooth type. Although no exact definition of a smooth alternative distribution has been made, it may be stated here that *smooth*, in the sense used by Neyman, will imply that the number of sets of signs will be small. For example, if the hypothesis tested is that observations follow a given normal curve, whereas in fact they have been drawn from a normal distribution identical with the first but with a smaller mean, then the differences between observation and expectation on the basis of the hypothesis tested may be expected to give a preponderance of positive signs below the sample mean and of negative signs above it; that is to say, if the difference in means is sufficient to offset the sampling fluctuations we should find a single set of positive signs followed by a single set of negative signs. If the true population is a normal curve with the same mean but with a larger standard deviation than that specified by the hypothesis tested, then there will be a tendency towards a set of positive signs, a set of negative signs, followed by a set of positive signs, although sampling fluctuations may not leave such a clear-cut answer. The more complex the alternative hypothesis the less chance there will be of detecting it.

7. With this objective in view it is proposed to take T, the number of sets of signs, as the test criterion, rejecting the hypothesis tested whenever, for a given r_1 and r_2, T is exceptionally small. This we do on the grounds that the existence of very few sets of signs suggests that the differences between observed and expected frequencies are not due to chance sampling fluctuations but to some systematic departure of the true probability law (assumed *smooth*) from hypothesis. In following this procedure we should reject the hypothesis if

$$P\{T \leqslant T_0\} = \sum_{T=2}^{T_0} P\{T \mid r_1 r_2\} \leqslant \epsilon,$$

where T_0 is the observed value of T and ϵ the significance level selected as appropriate. Exact probabilities are given in Table 1, and the application of the test is immediate.* There seems to be no reason why the test should not be applicable to both grouped and ungrouped observations, although the formulation of the hypothesis tested may be somewhat different in the two cases. Consider a sample which has been supposedly randomly

* An assumption implicit in the test would appear to be that for each χ^2 cell there is an equal chance of obtaining a positive or a negative deviation, that is, that there are sufficient numbers in each cell for the binomial to be closely approximated to by a normal curve. An extensive series of random sampling experiments has shown, however, that the divergence between theory and practice is not significant even when the probability of obtaining a positive is four times that of obtaining a negative. Hence while strictly the expectation in each cell of χ^2 should be 10 or over, it would seem that for practical purposes that the T test may be applied in all cases where the application of the χ^2 test is permissible.

Table 1. *Probability of obtaining a given number of sets, T.* [$T = 2t$ or $2t+1$]

The function tabled is $\dfrac{2(r_1-1)!\,(r_2-1)!}{(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!}$ or $\dfrac{(r_1-1)!\,(r_2-1)!\,(r_1+r_2-2t)}{t!(t-1)!\,(r_1-t)!\,(r_2-t)!}$, according as T is even or odd.

$P\{T\}$ is obtained by dividing this function by the binomial term $\dfrac{r!}{r_1!\,r_2!}$.

r	r_1	r_2	$\dfrac{r!}{r_1!\,r_2!}$	$T=2$	3	4	5	6	7	8	9	10	11	12	13	14
2	1	1	2	2												
3	2	1	3	2	1											
4	3	1	4	2	2											
	2	2	6	2	2	2										
5	4	1	5	2	3											
	3	2	10	2	3	4	1									
6	5	1	6	2	4											
	4	2	15	2	4	6	3									
	3	3	20	2	4	8	4	2								
7	6	1	7	2	5											
	5	2	21	2	5	8	6									
	4	3	35	2	5	12	9	6	1							
8	7	1	8	2	6											
	6	2	28	2	6	10	10									
	5	3	56	2	6	16	16	12	4							
	4	4	70	2	6	18	18	18	6	2						
9	8	1	9	2	7											
	7	2	36	2	7	12	15									
	6	3	84	2	7	20	25	20	10							
	5	4	126	2	7	24	30	36	18	8	1					
10	9	1	10	2	8											
	8	2	45	2	8	14	21									
	7	3	120	2	8	24	36	30	20							
	6	4	210	2	8	30	45	60	40	20	5					
	5	5	252	2	8	32	48	72	48	32	8	2				
11	10	1	11	2	9											
	9	2	55	2	9	16	28									
	8	3	165	2	9	28	49	42	35							
	7	4	330	2	9	36	63	90	75	40	15					
	6	5	462	2	9	40	70	120	100	80	30	10	1			
12	11	1	12	2	10											
	10	2	66	2	10	18	36									
	9	3	220	2	10	32	64	56	56							
	8	4	495	2	10	42	84	126	126	70	35					
	7	5	792	2	10	48	96	180	180	160	80	30	6			
	6	6	924	2	10	50	100	200	200	200	100	50	10	2		
13	12	1	13	2	11											
	11	2	78	2	11	20	45									
	10	3	286	2	11	36	81	72	84							
	9	4	715	2	11	48	108	168	196	112	70					
	8	5	1287	2	11	56	126	252	294	280	175	70	21			
	7	6	1716	2	11	60	135	300	350	400	250	150	45	12	1	
14	13	1	14	2	12											
	12	2	91	2	12	22	55									
	11	3	364	2	12	40	100	90	120							
	10	4	1001	2	12	54	135	216	288	168	126					
	9	5	2002	2	12	64	160	336	448	448	336	140	56			
	8	6	3003	2	12	70	175	420	560	700	525	350	140	42	7	
	7	7	3432	2	12	72	180	450	600	800	600	450	180	72	12	2

drawn from some population. Let the elements of the sample in order of drawing be $x_1, x_2, \ldots, x_n$. We may use the T criterion to test the hypothesis of randomness, in the following way. If u_1 is the smallest value observed in the sample and u_n the largest, then if we exclude the trivial case when all the x's are equal it is easy to show that $u_1 < \bar{x} < u_n$, where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$. If we now consider the deviations

$$x_i - \bar{x} = \delta x_i \quad \text{for} \quad i = 1, 2, \ldots, n,$$

there will be a series $\delta x_1, \delta x_2, \ldots, \delta x_n$, some of which quantities will be positive and some negative. The application of the T test is immediate, the admissible alternate hypotheses being that if the drawing of the sample is not at random then bias of the smooth kind is present.

8. As an illustration consider the following two cases:

Case I.	Expected frequency	10	25	35	75	155	155	75	35	25	10
	Observation	12	29	45	81	160	145	69	31	20	8
	Deviation	+	+	+	+	+	−	−	−	−	−
Case II.	Expected frequency	10	25	35	75	155	155	75	35	25	10
	Observation	12	23	45	66	161	160	69	36	20	8
	Deviation	+	−	+	−	+	+	−	+	−	−

In the first case $\chi^2 = 6{\cdot}94$ and in the second $\chi^2 = 6{\cdot}80$; in neither case would the hypothesis be rejected as inadequate by using the χ^2 criterion. The T criterion does, however, bring out the essential difference:

Case I. $\quad r_1 = 5, \quad r_2 = 5, \quad T_0 = 2 \quad$ and $\quad P\{T \leqslant T_0\} = \frac{2}{252}$.

Case II. $\quad r_1 = 5, \quad r_2 = 5, \quad T_0 = 8 \quad$ and $\quad P\{T \leqslant T_0\} = \frac{242}{252}$.

Using the T test we should be inclined to reject the first hypothesis in favour of a smooth *alternative, while for the second case we should be inclined to agree with the conclusion* drawn from the χ^2 test that the observational material is adequately described.

9. Sampling material is available whereby the theoretical distribution of T may be tested in practice. Neyman & Pearson (1928) took 208 samples, each of size 200, from a population of eight groups described by the cubic curve

$$y = 25 + \tfrac{45}{8}x - \tfrac{15}{128}x^3.$$

The expectation in each cell for a sample of this size was calculated and the χ^2 criterion found for each of the 208 samples. The writer was given access to these calculations and was able to find the sampling distribution of T from the material. The results of this sampling experiment and the theoretical distribution of T from relations (1) and (2) are given in Table 2.

The agreement between theory and practice would seem to be reasonably good, and in the cases (4, 4) and (5, 3) the values of χ^2, calculated to test the discrepancy between theory and practice, were not greater than might be attributable to sampling fluctuations. It was not thought worth while to calculate χ^2 for (6, 2) and (7, 1). A second sampling experiment in which samples of size 360 were drawn from a normal population of fifteen groups lent further support to the reasonableness of the theoretical distribution.

10. The T criterion will be a useful supplementary criterion to the χ^2, but because it takes account solely of the sign of a distribution and not of its magnitude it will probably only be useful when used in conjunction with χ^2. A test of significance which could combine both the probability levels of T and χ^2 would undoubtedly be more useful, and we may

therefore consider how this might be done. Unless the exact degree of dependence which exists between two variables is known it is usually only possible to obtain their joint distribution if they are independent. It would appear reasonable, both on theoretical grounds and from sampling experiments, to assume that T and χ^2 are independent, or, if the assumptions underlying both tests are not exactly fulfilled, to assume that the degree of dependence between them is at most small.

Table 2. *Comparison of theoretical distribution of T with that derived from a sampling experiment*

(4 positive, 4 negative)

T = number of sets	2	3	4	5	6	7	8	Total
Sampling	3	5	20	25	28	5	7	93
Theory	2·7	8·0	23·9	23·9	23·9	8·0	2·7	93

(5 positive, 3 negative) or (3 positive, 5 negative)

T = number of sets	2	3	4	5	6	7	Total
Sampling	2	8	32	30	20	10	102
Theory	3·6	10·9	29·1	29·1	21·9	7·3	102

(6 positive, 2 negative) or (2 positive, 6 negative)

T = number of sets	2	3	4	5	Total
Sampling	1	3	—	5	9
Theory	0·6	1·9	3·2	3·2	9

(7 positive, 1 negative) or (1 positive, 7 negative)

T = number of sets	2	3	Total
Sampling	—	2	2
Theory	0·5	1·5	2

11. We shall begin by demonstrating that as far as mathematics are concerned the T and χ^2 criteria are completely independent.* For simplicity of argument let us consider the case of three groups only. The sample may then be represented by a point (n_1, n_2, n_3) in three-dimensioned space, with axes of reference On_1, On_2, On_3, and the expected population values by a point (m_1, m_2, m_3) in the same space. Since

$$n_1+n_2+n_3 = m_1+m_2+m_3 = N,$$

* This method of approach was suggested to me by Andrew Gleason of Harvard University at a seminar given at the Statistical Laboratory, University of California, at Berkeley.

these points are constrained to lie in a plane. Fig. 1 shows this plane for the particular case $N = 16; m_1 = 4, m_2 = 8, m_3 = 4$. Since no frequency can be negative, possible sample points must be within an equilateral triangle lying in this plane, the chance of occurrence associated with a point being the multinomial term

$$\frac{N!}{n_1!\,n_2!\,n_3!}\left(\frac{m_1}{N}\right)^{n_1}\left(\frac{m_2}{N}\right)^{n_2}\left(\frac{m_3}{N}\right)^{n_3}.$$

When using the χ^2 test the mathematical approximation consists in substituting for this term an expression proportional to $e^{-\frac{1}{2}\chi^2}$, in regarding this last as a continuous function, and

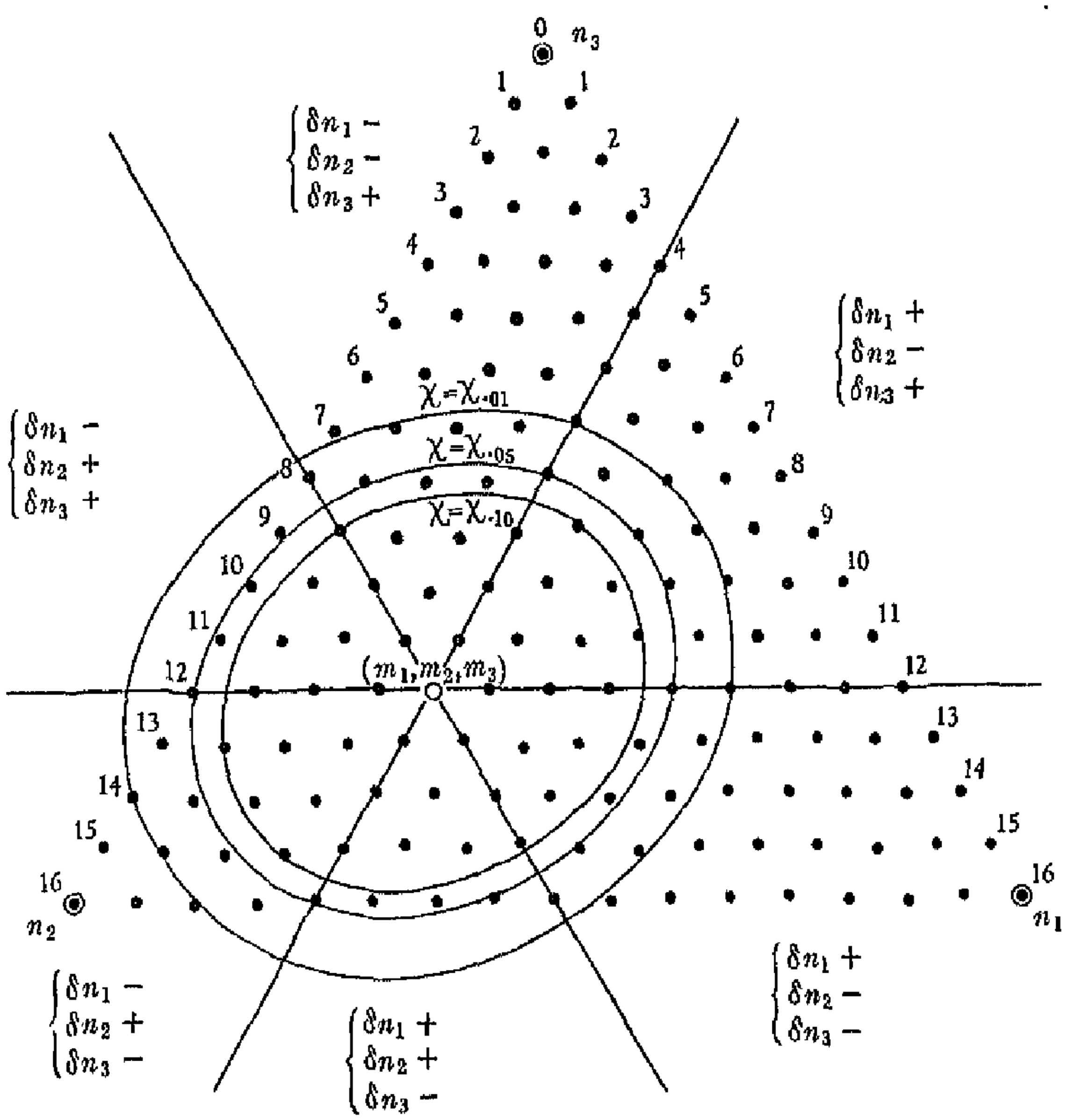


Fig. 1. Graphical illustration of the χ^2 contours and the change in signs of the δn's. n_1, n_2 and n_3 denote the points of intersection of the On_1, On_2, On_3 axes with the plane $n_1+n_2+n_3 = N$. According to the approximation, the chance equals α of obtaining a sample point lying outside the elliptic contour on which $\chi = \chi_\alpha$.

in taking as a measure of goodness of fit the integral of this expression outside the ellipse which passes through the sample point and on which χ^2 is constant. For the case of three *groups this integral itself assumes the simple form $e^{-\frac{1}{2}\chi^2}$. Three such elliptic contours are* shown in the diagram.

Planes through (m_1, m_2, m_3) parallel to the co-ordinate planes n_1On_2, n_2On_3, n_3On_1, will intersect the sample plane

$$n_1+n_2+n_3 = N$$

in three straight lines. As shown in the diagram, these lines divide the sample plane into six sectors, and for all sample points within a sector the signs of the differences $\delta n_i = n_i - m_i$ will remain unchanged. Any test based solely on runs of signs will consist in taking one or more of these sectors as critical regions and rejecting the hypothesis tested when the sample point falls therein. It is clear that if we use the mathematical approximation, the distribution of χ^2 is the same within each sector; similarly, that the chance of a sample showing a given combination of signs is the same on each ellipse along which χ^2 is constant. Thus under the assumptions made regarding the distribution of χ^2, the T and χ^2 criteria are completely independent.

In this case of three groups T can only assume values of two or three and the former value would not be judged significant, but the argument will follow exactly similar lines in the case of many groups. The number of sectors will be in general $2(2r_2+1)$ if $r_1 > r_2$ and $2(2r_2)$ if $r_1 = r_2$, and they will be bounded by primes passing through the population point.

12. While the distributions of T and χ^2 are independent for this mathematical model they are unlikely to be exactly so when we go back to the true multinomial density distribution, because the sample space is neither continuous nor infinite. The model, in fact, becomes inaccurate if m_1, m_2 or m_3 are very small. For example, it is seen in Fig. 1 that while the 1 % ellipse ($\chi = \chi_{0 \cdot 01}$) lies completely within the triangular space for the sectors with signs $+ - +$ and $- - +$, it lies completely without the space for the sector $+ + -$ and partly without for the other sectors. It has been thought worth while therefore to test whether the two criteria are independent in practice, and to this end the same material previously described has been utilized. Tables 3 and 4 give the distribution of mean χ^2 for different values of $P\{T\}$ and the distribution of mean $P\{T\}$ for grouped values of χ^2. There is little evidence in these figures to show that $P\{T\}$ and χ^2 (and therefore $P\{\chi^2\}$) are related. The figures therefore lend support to the geometrical argument and indicate that the approximations involved in χ^2, both from the small sample and the fact that the sample space is not infinite, do not invalidate the mathematical result.

13. In order to combine the χ^2 and T tests of significance it will be necessary to develop a theory for the combination of two tests of significance when one criterion is a continuous and the other a discontinuous variable. R. A. Fisher has set out the test for the combination of tests of significance from a number of independent continuous variables. The keystone of the test is the recognition of the fact that if Z is a continuous variable, then z, where

$$z = \int_{-\infty}^{Z} p(Z)\,dZ,$$

is also a continuous variable equally likely to have any value between 0 and 1; we shall describe z as being distributed rectangularly. Twice the logarithm of the product of two such z's, say z_1 and z_2, where z_1 and z_2 follow from two independent tests of significance can be shown to be distributed as χ^2 with four degrees of freedom. Consider a discontinuous variable X which may take values $X_1, X_2, \ldots, X_s$ and which has an elementary probability law $P\{X = X_j\} = p_j$, where

$$0 < p_j < 1 \quad \text{for} \quad j = 1, 2, \ldots, s$$

and

$$\sum_{j=1}^{s} p_j = 1.$$

If a new variable, x, is defined as taking values $x_1, x_2, \ldots, x_s$, where

$$x_k = \sum_{j=1}^{k} p_j,$$

Table 3. *Mean χ^2 for different values of $P\{T\}$*

r_1, r_2 or r_2, r_1	—	4, 4	5, 3	4, 4	5, 3	6, 2	4, 4	5, 3	4, 4	6, 2
No. of obs. on which mean is based	26	5	20	28	30	—	25	32	20	3
$P\{T\}$	1·00	0·97	0·93	0·86	0·71	0·64	0·63	0·43	0·37	0·29
Mean χ^2	7·19	6·04	5·87	6·43	5·98	—	7·41	7·32	6·82	4·86

r_1, r_2 or r_2, r_1	7,1	5, 3	4, 4	6, 2	5, 3 and 4, 4
No. of obs. on which mean is based	—	8	5	1	5
$P\{T\}$	0·25	0·14	0·11	0·07	0·04 and 0·03
Mean χ^2	—	6·10	5·53	6·95	8·06

Table 4. *Mean $P\{T\}$ for grouped χ^2*

χ^2	0·0–1·0	1·0–2·0	2·0–3·0	3·0–4·0	4·0–5·0	5·0–6·0	6·0–7·0	7·0–8·0	8·0–9·0	9·0–10·0
No. of obs. on which mean is based	1	7	14	27	32	29	21	14	11	15
Mean $P\{T\}$	0·71	0·73	0·69	0·60	0·68	0·60	0·65	0·57	0·67	0·64

χ^2	10·0–11·0	11·0–12·0	12·0–13·0	13·0–14·0	14·0–15·0	15·0–16·0	16·0–17·0	17·0–18·0
No. of obs. on which mean is based	10	10	5	4	2	1	2	1
Mean $P\{T\}$	0·74	0·65	0·74	0·58	0·82	1·00	0·63	0·63

then x_k *may only take values between* 0 *and* 1 *for* $k = 1, 2, \ldots, s$. It is required to find the joint probability law of the product of two independent variables x and z, where x and z are as defined above. It will be noted that the elementary probability law of x will be

$$P\{x = x_j\} = p_j \quad (j = 1, 2, \ldots, s).$$

Hence when $x = x_j$ (the probability of which is p_j), the product xz will be distributed rectangularly between 0 and x_j on a proportion p_j of occasions. It follows that xz has a probability distribution which has points of discontinuity at $x_1, x_2, \ldots, x_s$, that it is distributed rectangularly between these points of discontinuity, and that

$$P\{0 < xz < x_1\} = p_1 \sum_{j=1}^{s} \frac{p_j}{x_j}, \quad P\{x_1 < xz < x_2\} = p_2 \sum_{j=2}^{s} \frac{p_j}{x_j}.$$

Generally

$$P\{x_{k-1} < xz < x_k\} = p_k \sum_{j=k}^{s} \frac{p_j}{x_j}.$$

14. If we now apply this theory to the combination of the tests of significance of T and χ^2, it is seen that we must consider the product of $P\{\chi^2\}$ and $P\{T\}$. χ^2 is a continuous variable and

$$z = \int_{\chi_0^2}^{+\infty} p(\chi^2)\, d(\chi^2) = P\{\chi^2 \geqslant \chi_0^2\} = P\{\chi^2\}$$

is distributed rectangularly between 0 and 1, and

$$x = \sum_{T=2}^{T_0} P\{T \mid r_1 r_2\} = P\{T \leqslant T_0\} = P\{T\}$$

is a discontinuous variable taking known values. The probability integral of xz is thus known from theory and $Y_{0\cdot05}$ or $Y_{0\cdot01}$ can be found to satisfy the relation

$$P\{0 < xz < Y_\epsilon\} = \epsilon.$$

These probability levels are given in Table 5. The procedure for the joint test of significance will be:

(i) *calculate $P\{T\}$ as described in § 7;*

(ii) calculate $P\{\chi^2\}$ in the usual way. The degrees of freedom will be the number of groups minus one;

(iii) multiply $P\{T\}$ and $P\{\chi^2\}$ together and refer to Table 5 to judge the significance of the product.

Table 5. *Values of $Y_{0\cdot05}$ and $Y_{0\cdot01}$, where $P\{P(\chi^2)\,P(T) < Y_\epsilon\} = \epsilon$*

This table may be used to judge the significance of the joint distribution of the T criterion and any other continuous criterion.

r	r_1	r_2	$Y_{0\cdot05}$	$Y_{0\cdot01}$	r	r_1	r_2	$Y_{0\cdot05}$	$Y_{0\cdot01}$
5	4	1	0·0312^5	0·0062^5	11	10	1	0·0275+	0·0055+
	3	2	0·0213	0·0043		9	2	0·0171	0·0034
						8	3	0·0144	0·0028
6	5	1	0·0300	0·0060		7	4	0·0144	0·0025+
	4	2	0·0211	0·0042		6	5	0·0140	0·0024
	3	3	0·0195	0·0039					
					12	11	1	0·0273	0·0055−
7	6	1	0·0292	0·0058		10	2	0·0174	0·0035−
	5	2	0·0197	0·0039^5		9	3	0·0149	0·0027
	4	3	0·0174	0·0035		8	4	0·0142	0·0024
						7	5	0·0135	0·0022
8	7	1	0·0286	0·0057		6	6	0·0131	0·0021
	6	2	0·0188	0·0038					
	5	3	0·0160	0·0032	13	12	1	0·0271	0·0054
	4	4	0·0153	0·0031		11	2	0·0165+	0·0033
						10	3	0·0151	0·0026
9	8	1	0·0281	0·0056		9	4	0·0138	0·0023
	7	2	0·0180	0·0036		8	5	0·0137	0·0022
	6	3	0·0153	0·0031		7	6	0·0138	0·0022^5
	5	4	0·0140	0·0028					
					14	13	1	0·0269	0·0054
10	9	1	0·0278	0·0056		12	2	0·0163	0·0033
	8	2	0·0175	0·0035		11	3	0·0151	0·0025+
	7	3	0·0143	0·0029		10	4	0·0135−	0·0022^5
	6	4	0·0143	0·0026		9	5	0·0138	0·0023
	5	5	0·0143	0·0025		8	6	0·0136	0·0022
						7	7	0·0134	0·0022

15. The application of the joint test of significance may be illustrated by means of an example. A sample of 360 observations is available. This sample has actually been randomly drawn from a normal population of which the mean is zero and the standard deviation unity. The figures are given in Table 6. Calculations give $\chi^2 = 21\cdot 1$ and $P\{\chi^2\} = 0\cdot 10$. Judging by the χ^2 alone we should say probably that there is nothing out of the ordinary in the deviations of the sample from the expected values. The number of signs is 15, of which 9 are positive and 6 negative, and these are arranged in six sets. Making the appropriate calculations, we have

$$P\{6 \text{ sets} \mid 9 \text{ positive}; 6 \text{ negative}\} = \tfrac{875}{5005} = 0\cdot 175.$$

The arrangement of signs will therefore be judged as acceptable. The joint significance of a $P\{\chi^2\} = 0\cdot 10$ and a $P\{T\} = 0\cdot 175$ is found, by evaluating the joint distribution, to be 0·066.

Table 6. *Sample values. Observed and expected*

Central values	−2·1 and under	−1·8	−1·5	−1·2	−0·9	−0·6	−0·3	0·0
Observation Expectation	12 9·3	10 8·6	18 14·0+	26 21·0+	23 28·7	42 35·9	43 41·0+	49 43·0−
Deviation	+2·7	+1·4	+4·0	+5·0	−5·7	+6·1	+2·0	+6·0

Central values	+0·3	+0·6	+0·9	+1·2	+1·5	+1·8	+2·1 and over	Total
Observation Expectation	35 41·0+	28 35·9	20 28·7	26 21·0	20 14·0	3 8·6	5 9·3	360 360
Deviation	−6·0	−7·9	−8·7	+5·0	+6·0	−5·6	−4·3	0

16. A study of the basic table (Table 1) of the function T will show that $P\{T\}$ is not a very sensitive criterion with which to judge the randomness of a sequence of signs unless the number of groups under consideration is very large. For example, if there are 10 signs, 5 of which are positive and 5 of which are negative, the probability of getting two sets of signs is 0·008. Thus the test would show, and rightly, that the chance of such an arrangement is small, but this fact would undoubtedly be recognized by a skilled computer without the use of a test at all. In the case of 10 signs the probability of three groups or less is 0·040, and this would possibly be judged non-significant. Again, let us consider an extreme case say, 10 signs, 9 of which are positive and 1 negative. The T criterion does not concern itself with the fact that the numbers 9 and 1 are exceptional, it is merely concerned with deciding whether their arrangement is exceptional given the 9 and 1. Table 1 shows that neither possible arrangement would be considered out of the ordinary. It is these points of weakness which show that the criterion T is not of great utility except in combination with χ^2. For, if we consider the 9 positive, 1 negative case, common sense tells us that the χ^2 criterion in such

a case would possibly be significant. Nine positive deviations have to be balanced by a single negative deviation, and this last is therefore likely to be big. This does not influence T; neither will the contribution of T to the joint criterion be of much weight. This is as it should be, for it is difficult to see how one can postulate a smooth alternative for 9 positive, 1 negative, two sets, and not also for 9 positive, 1 negative, three sets. Generally, however, we shall not meet such extreme cases in practice. One way of overcoming this weakness of the test would be to consider the probability of obtaining r_1 positive and r_2 negative signs together with the probability of obtaining T sets of alternate positive and negative signs given r_1 and r_2. This is simple enough when considering just a sequence of alternatives, as I have shown elsewhere, but it is not easy to fit these results to the χ^2 problem, nor, when this is possible, will the choice of a critical region be straightforward. However, the results of sampling experiments will be utilized to throw light on these points and it is hoped to discuss them, with other questions arising, in a further publication.

17. It is possible that there are other criteria, depending on the arrangement of positive and negative signs, which will be more sensitive than the T criterion chosen. For example, it is easy to calculate, given r_1 and r_2, the probability that the largest set is composed of a sequence of r' positive signs, and there are many other possibilities which might be considered. It would appear that any criterion based on sign sequences can be shown to be independent of χ^2 by means of geometrical argument, and it will be necessary therefore to consider the power of these different sign tests when referred to a specified set of alternate hypotheses.

18. The main objection to the two criteria, T and $P\{\chi^2\}.P\{T\}$, that I have proposed in this note is the one which was mentioned earlier; they are only applicable to the case where there is just one restriction on χ^2, i.e. when the totals of expected and observed frequencies have been made to agree. It is possible to work out a slightly different form of the T criterion for each additional restriction which is put on χ^2, and this has been done. It is preferable, however, to delay publication until the results of an extensive sampling experiment are complete in order to verify whether such theoretical assumptions as have been made are reasonable.

REFERENCES

Neyman, J. (1937). *Skand. Aktuar. Tidskr.* **20**, 149–99.

Neyman, J. & Pearson, E. S. (1928). *Biometrika*, **20** A, 263–94.

AN EXACT TEST FOR THE EQUALITY OF VARIANCES*

By R. L. PLACKETT, M.A.

Introduction

The problem of testing the equality of variances and covariances in normal distributions is one which has received considerable attention; we have compiled a bibliography of some sixty papers, and shall issue a survey of these in due course; only papers vital to our discussion will be considered here. A precise instance of the type of situation we are considering is as follows: measurements of height, span and tibia length are made on each of 20 Englishmen, 20 Scotsmen, 20 Welshmen and 20 Irishmen; it is required to know if the covariance matrix of the three characteristics is the same for each of the four nationalities. Nothing is known or assumed about the mean values of these characteristics in the four populations considered, nor are we interested in testing any hypothesis concerning the means, although such a hypothesis may be the object of further investigations which assume that the four covariance matrices are the same; this latter assumption is inevitably made in multivariate analysis of variance.

Wilks (1932) has already given the moments of the distribution of his criterion for testing the equality of several covariance matrices (on the hypothesis that the matrices are in fact equal) and Bishop (1939) put this criterion into an approximate workable shape. The test criterion given here differs from that of Wilks and has the advantage when one or two correlated characteristics are being measured (height or height and span, for example) that its distribution is exactly known whatever the number of populations. Nair (1939) did, it is true, give the exact distribution of the Neyman & Pearson (1931) L_1 criterion for one measured characteristic; and the exact distribution for two characteristics of Wilks's generalization of their criterion; but the form in which the distribution was obtained is very involved. It is interesting to notice that from our standpoint the problem of testing the equality of several variances (i.e. the case of one measured characteristic) is, as will appear, brought within the framework of multiple correlation theory. In the general case of more than two characteristics the moments of the distribution of our criterion, like those of Wilks, are available.

Outline of method

In the usual terminology we consider k p-variate normal distributions and are concerned with testing the hypothesis that the corresponding variances and covariances are all equal. The method we employ to test this hypothesis is essentially that which has been in use in analysis of variance since its origination by Fisher; to test the equality of a set of k quantities we test whether $(k-1)$ orthogonal linear functions of the quantities are each zero. To illustrate the application of this principle in the present instance take the particular case $p = 1$, i.e. we wish to test the equality of the variances in k univariate normal distributions. If a typical observation from the lth distribution is t_l $(l = 1, 2, \ldots, k)$, form k mutually orthogonal linear functions of the t_l such that one is

$$u = t_1 + t_2 + \ldots + t_k.$$

* Communication from the National Physical Laboratory.

If the $(k-1)$ covariances of u and each of the other linear functions are all zero then the variances of the k distributions must all be equal; this condition may be expressed by saying that the multiple correlation coefficient of u on the other linear functions is zero. Further, if there are n sample values of u then the size of sample drawn from each distribution must also be n at least, and if no observations are to be discarded the size of each sample must be n exactly. Thus, although it is not a condition of the problem that the sizes of samples drawn from the k distributions must all be equal, it is a condition of our solution.

The extension of the foregoing principle to $p > 1$ is straightforward and is considered in detail in the next section; the problem then becomes that of testing the independence of two groups of variates, the first of size p, i.e. p expressions of the form u; and the second of size $p(k-1)$ comprising all the other orthogonal linear functions. This problem has been treated by Wilks (1935, 1943) and the relevant distribution is expressible as an incomplete β-function when $p = 1$ and 2 (for all k); an exact distribution is also known when $p = 3$ and 4 for $k = 2$. Finally, since when $p = 1$ the criterion has the form of a multiple correlation coefficient, the power of the test in this instance can be calculated by virtue of the work of Fisher (1928).

Discussion of the test

A sample of n observations is drawn from each of the k p-variate normal distributions of which the lth has the covariance matrix V^l_{ij} $(l = 1, 2, \ldots, k;\ i, j = 1, 2, \ldots, p)$. It is required to test the hypothesis that

$$V^l_{ij} = V^m_{ij} \quad (l, m = 1, 2, \ldots, k). \tag{1}$$

The population means do not enter into the hypothesis and have arbitrary unknown values. Where i, j, l, m appear henceforth they will be understood to range over the values given above unless otherwise stated. The observations may be written in the form of an $n \times kp$ matrix X such that all those on the ith variate in the lth distribution are in column $(i-1)k + l$. The αth observation in this column $(\alpha = 1, 2, \ldots, n)$ is denoted by $x^l_{i\alpha}$; the order of the elements in a column is assumed to be random. If this is doubted the observations should be randomly rearranged.

We must emphasize here that the sample value of the criterion to be used to test (1) depends on this order, and there is thus, in a sense, a correspondence between $x^l_{i\alpha}$ and $x^m_{j\alpha}$, although these two quantities are, of course, uncorrelated when $l \neq m$. Most tests of a hypothesis specifying nothing about the order in which observations are made or written down are themselves independent of it; ours is not, and different computers with the same data might well come to different conclusions although this does not affect the validity of the test, the significance level being overall what it should be. There is probably some loss of power which can, however, be offset by imbuing α with a certain physical meaning; but we shall not discuss this question here. A criterion for testing normality depending on the order of arrangement of observations has been suggested by R. C. Geary (1935, pp. 316–17).

Let now
$$z^l_{i\alpha} = x^l_{i\alpha} - \left(\sum_\alpha x^l_{i\alpha}\right) \Big/ n, \tag{2}$$

and let the corresponding $n \times kp$ matrix be Z. If $G = Z'Z$, where a prime is used to denote the transpose of a matrix, then, apart from a factor n, G is the matrix of sample variances and covariances of all variables. We further define $S(k, p)$ as the sum of all (k^{2p}) signed minors

$$g^{m_1, m_2, \ldots, m_p}_{l_1, l_2, \ldots, l_p},$$

formed by rows $l_1, l_2, \ldots, l_p$ and columns $m_1, m_2, \ldots, m_p$ of G, where

$$(i-1)\,k < l_i,\ m_i \leqslant ik. \tag{3}$$

$\tilde{S}(k,p)$ is similarly defined for the matrix $\tilde{G} = G^{-1}$ (we shall use this notation for the inverses of matrices throughout).

We now proceed to prove the following

THEOREM: $$W(k,p) = k^{2p}/S(k,p)\,\tilde{S}(k,p)$$

is distributed like Wilks's statistic for testing the hypothesis that two groups of variates of sizes p and $p(k-1)$, known to have been drawn from a (kp)-variate normal distribution, are mutually independent (Wilks, 1935, 1943). If the groups are in fact mutually independent then (1) is true.

Proof. Introduce a $k\times k$ orthogonal matrix B, the elements of whose first column are all equal (to $\pm 1/\sqrt{k}$) but which is otherwise quite arbitrary. Put

$$r = (i-1)\,k+l, \quad u = (m-1)\,p+j, \tag{4}$$

and form a $kp\times kp$ matrix A such that

$$a_{ru} = \delta_{ij}\,b_{lm}, \tag{5}$$

where $\delta_{ij} = 1$ $(i=j)$, otherwise O. Clearly A is also orthogonal. For example, suppose $k=4$, $p=2$. Apart from a factor of $\pm\frac{1}{2}$ multiplying each element, let

$$B = \begin{Vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{Vmatrix}.$$

Then $$A = \begin{Vmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \end{Vmatrix}.$$

When $p=1$, $A=B$. Let

$$D = XA, \quad Y = ZA, \quad C = Y'Y = A'GA. \tag{6}$$

Putting $$s = (j-1)\,k+m, \quad t = (l-1)\,p+i, \tag{7}$$

and defining $$t' = (l-1)\,p+i, \quad u' = (m-1)\,p+j \quad (l,m = 2,3,\ldots,k), \tag{8}$$

we have $$\mathscr{E}(g_{rs}) = \delta_{lm}(n-1)\,V^l_{ij}, \tag{9}$$

so that $$\mathscr{E}(c_{iu}) = (n-1)\sum_l b_{lm} V^l_{ij}/\sqrt{k}. \tag{10}$$

Hence when $$\mathscr{E}(c_{iu'}) = 0 \tag{11}$$

equations (1) are satisfied, because for fixed i and j equations (10) can be solved and yield

$$(n-1)\,V^l_{ij} = \mathscr{E}(c_{ij}). \tag{12}$$

Denote a typical element of the tth column of D by d_t. Then equations (11) are satisfied if and only if d_t and $d_{u'}$ are mutually independent.

A criterion for testing (11), obtained by likelihood-ratio methods, has been given by Wilks (1935, 1943). This is

$$W(k,p) = \frac{|c_{tu}|}{|c_{ij}|\,|c_{t'u'}|}, \tag{13}$$

and is sometimes called the vector alienation coefficient. Let $C^{(p)}$ be the pth compound of C (Aitken, 1939, p. 90), i.e. the matrix of all $p \times p$ minors of C; and $\tilde{C}^{(p)}$ the pth compound of $\tilde{C} = C^{-1}$ (since $\tilde{C}^{(p)}$ is the inverse of $C^{(p)}$ our notation is consistent). Then

$$W(k,p) = 1/c_{11}^{(p)}\tilde{c}_{11}^{(p)} \tag{14}$$

by an application of Jacobi's theorem on the minors of the adjugate (Aitken, 1939, p. 97). Now by the Binet-Cauchy theorem (Aitken, 1939, p. 93),

$$C^{(p)} = (A')^{(p)}\,G^{(p)}\,A^{(p)}, \quad \tilde{C}^{(p)} = (A')^{(p)}\,\tilde{G}^{(p)}\,A^{(p)}. \tag{15}$$

Consider the elements in the first row of $(A')^{(p)}$. The first p rows of A' are of the form

$$\begin{Vmatrix} 11\ldots1 & 00\ldots0 & 00\ldots0 & \ldots & 00\ldots0 \\ 00\ldots0 & 11\ldots1 & 00\ldots0 & \ldots & 00\ldots0 \\ 00\ldots0 & 00\ldots0 & 11\ldots1 & \ldots & 00\ldots0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 00\ldots0 & 00\ldots0 & 00\ldots0 & \ldots & 11\ldots1 \end{Vmatrix},$$

apart from the factor $\pm 1/\sqrt{k}$ multiplying each element. Therefore the only non-zero elements in the first row of $(A')^{(p)}$ are those formed by taking one column from each of the p blocks of k columns into which the first p rows of A' may be divided. All the non-zero elements equal $k^{-\frac{1}{2}p}$. Then from (15)

$$c_{11}^{(p)} = S(k,p)\,k^{-p}, \quad \tilde{c}_{11}^{(p)} = \tilde{S}(k,p)\,k^{-p}, \tag{16}$$

so finally

$$W(k,p) = k^{2p}/S(k,p)\,\tilde{S}(k,p). \tag{17}$$

This completes the proof.

Case of $p = 1$

Here
$$W(k,1) = k^2/S(k,1)\,\tilde{S}(k,1),$$

where $S(k,1)$, $\tilde{S}(k,1)$ are the sums of all elements of G, G^{-1} respectively. If (1) is true, $\mathbf{W}(k,1)$, the true value of $W(k,1)$, is unity. Define

$$W(k,1) = 1 - R^2 \quad \text{and} \quad \mathbf{W}(k,1) = 1 - \mathbf{R}^2, \tag{18}$$

so that if (1) is true, $\mathbf{R} = 0$. The distribution of $R^2 = 1 - W(k,1)$ when $\mathbf{R} = 0$ is, as Wilks pointed out, well known, being that of the multiple correlation coefficient (of d_1 on $d_2, d_3, \ldots, d_k$); if in the usual notation

$$I_x(a,b) = [B(a,b)]^{-1}\int_0^x x^{a-1}(1-x)^{b-1}\,dx, \tag{19}$$

then the cumulative distribution function of $x = R^2$ is $I_x(k-1, n-k)$, values near 1 being significant; that of $x = W(k,1)$ being $I_x(n-k, k-1)$ with small values significant. Tables in convenient form have been calculated by Thompson (1941); otherwise we can convert to the variance-ratio F by

$$F = (n-k)(1-W)/(k-1)\,W. \tag{20}$$

It is clear that n must exceed k; for p variates, n exceeds pk in order that G may be non-singular.

If the matrix A is defined instead as a $kp \times kp$ orthogonal matrix, the elements of whose first column are all equal (cf. equation (5)), the problem is effectively reduced to the case $p = 1$ whatever the value of p, and we can test exactly the somewhat indefinite hypotheses

$$V_{i1}^{l} + V_{i2}^{l} + \ldots + V_{ip}^{l} = V_{j1}^{m} + V_{j2}^{m} + \ldots + V_{jp}^{m}. \tag{21}$$

This may be applied in the following manner, for take $k = p = 2$ and obtain

$$V_{11}^{1} + V_{12}^{1} = V_{21}^{1} + V_{22}^{1} = V_{11}^{2} + V_{12}^{2} = V_{21}^{2} + V_{22}^{2}. \tag{22}$$

Thus
$$V_{11}^{1} = V_{22}^{1} \quad \text{and} \quad V_{11}^{2} = V_{22}^{2}. \tag{23}$$

If it is assumed
$$V_{11}^{1} = V_{11}^{2}, \tag{24}$$

then
$$V_{12}^{1} = V_{12}^{2}, \tag{25}$$

and conversely.

Case of $p = 2$

The distribution of $W(k, 2)$ has been given by Wilks (1935). If $x = \sqrt{[W(k, 2)]}$, the cumulative distribution function of x is

$$I_x(n-2k, 2k-2). \tag{26}$$

Small values of x are significant and n must exceed $2k$.

Case of $p \geqslant 3$

For $k = 2$, $p = 3$ and 4, the exact distributions are again known and have been given by Wilks in equations (35) and (37) respectively of his 1935 paper. The expressions are rather complicated and we have not reproduced them here. For other values of k and p the moments of $W(k, p)$ are available; while more recently Wald & Brookner (1941) have obtained the distribution in the form of an infinite series, calculating numerical values for the coefficients in certain instances.

For $p > 1$, (17) becomes rather intractable as a means of calculating $W(k, p)$. Indeed, for $k = 2$ and $p = 4$ it is necessary

(i) to calculate 36 sample variances and covariances,

(ii) find the inverse of an 8×8 matrix,

(iii) calculate 512 4×4 determinants,

and it is clearly better to reintroduce the matrix A in some appropriate numerical form, calculate $Y = ZA$ and $C = Y'Y$, and find $W(2, 4)$ from (13), a process which involves the evaluation of an 8×8 and two 4×4 determinants.

POWER OF THE TEST WHEN $p = 1$

From (17) the true value of R^2 is in general given by

$$1 - \mathbf{R}^2 = k^2 \bigg/ \left[\left(\sum_l V_{11}^l\right)\left(\sum_l 1/V_{11}^l\right)\right], \tag{27}$$

and thus the test will have equal power for all values of the variances such that the product of their sum and the sum of their reciprocals is constant. Consequently $1 - W(k, 1)$ is distributed like the multiple-correlation coefficient in samples from a population where the

true value is given by (27). The probability density function of this distribution has been deduced by Fisher (1928) and can be integrated to give a finite series when $(n-k)$ is even.

We find easily when $k = 2$ that in the V_{11}^1, V_{11}^2 quarter-plane the equipotentials are pairs of lines

$$V_{11}^1 = aV_{11}^2 \quad \text{and} \quad aV_{11}^1 = V_{11}^2, \tag{28}$$

where

$$a = (1+\mathbf{R})/(1-\mathbf{R}). \tag{29}$$

For $k > 2$ the equipotential surfaces in k dimensions are cones through the origin situated symmetrically with regard to the co-ordinate primes.

Reverting to $k = 2$ three methods are available for testing the hypothesis that $V_{11}^1 = V_{11}^2$:

(i) Fisher's z or $F = \exp(2z)$

$$= g_{11}/g_{22}. \tag{30}$$

(ii) the L_1 criterion introduced by Pearson & Neyman (1930) and later extended to $k > 2$ (Neyman & Pearson, 1931).

In the instance we are considering, i.e. equal sample sizes from both populations,

$$L_1 = 2(g_{11}g_{22})^{\frac{1}{2}}/(g_{11}+g_{22}) \tag{31}$$

$$= 2F^{\frac{1}{2}}/(1+F). \tag{31a}$$

(iii) $$W(2,1) = 4[g_{11}g_{22}-(g_{12})^2]/[(g_{11}+g_{22})^2-(2g_{12})^2].* \tag{32}$$

Thus tests (i) and (ii) are exactly equivalent, as is known, the optimum critical region being that corresponding to equal tails of the F-distribution. Criterion (iii) is that obtained by Morgan (1939) and Pitman (1939), appearing as equation (12) in Morgan's paper, to test that the variances in a normal bivarate population are equal. Morgan has compared the powers of tests (i) and (iii) for $n = 12$, 25 and 100 at a significance level of 0·10 and for these sample sizes it appears that the tests are effectively of equal power.

When n is large and, consequently, the two populations being independent, $(g_{12})^2/g_{11}g_{22}$ is converging in probability to zero,

$$W(2,1) \sim L_1^2. \tag{33}$$

The cumulative distribution functions of criteria (ii) and (iii) are respectively $I_x\left(\frac{n-1}{2}, \frac{1}{2}\right)$ (Nayer, 1936) $(x = L_1^2)$ and $I_x\left(\frac{n-2}{2}, \frac{1}{2}\right)$ $(x = W(2,1))$. Generally, $W(k,1)$ for large n converges in probability to the harmonic mean of the sample variances divided by their arithmetic mean; L_1 (for equal sample sizes) is exactly equal to the geometric mean divided by the arithmetic mean.

Example of the use of the test for a case with $k=4$, $p=1$

It is not easy to calculate $W(k,1)$ from equation (17) if $k > 3$. Indeed, the main value of (17) lies in showing the form of solution, and in establishing that this is independent of the particular orthogonal transformations used. In the following example, therefore, orthogonal transformations are made at once and the multiple correlation coefficient is calculated from the numerical data. This procedure is far quicker than that involved in calculating $W(4,1)$ from (17).

* See Appendix.

Below are given samples of 10 from each of four univariate normal populations:

x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
−20	+24	+ 4	+52	+ 7	+15	+ 8	− 8
− 1	+18	+ 9	−24	+ 5	+24	− 1	+56
−11	+27	−27	0	+18	−12	+ 1	−64
+10	+21	+ 5	+48	+13	−24	− 4	+12
− 4	−48	− 3	+48	− 6	+12	+ 5	−12

which have mean zero and standard deviations respectively 10, 30, 10, 40. Make the following orthogonal transformation:

$$y_1 = x_1+x_2+x_3+x_4, \quad y_2 = x_1+x_2-x_3-x_4,$$
$$y_3 = x_1-x_2+x_3-x_4, \quad y_4 = x_1-x_2-x_3+x_4,$$

and obtain

y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4
+60	−52	−92	+ 4	+22	+22	+ 8	−24
+ 2	+32	+14	−52	+84	−26	−76	+38
−11	+43	−65	−11	−57	+69	+95	−57
+84	−22	−54	+32	− 3	−19	+21	+53
− 7	−97	− 7	+95	− 1	+13	− 1	−35

Form the matrix of sums of squares and cross-products, i.e. C. This is

$$\begin{Vmatrix} +18636{\cdot}1 & -9646{\cdot}9 & -18232{\cdot}9 & +7325{\cdot}1 \\ -9646{\cdot}9 & +21784{\cdot}1 & +12018{\cdot}1 & -19015{\cdot}9 \\ -18232{\cdot}9 & +12018{\cdot}1 & +28692{\cdot}1 & -9445{\cdot}9 \\ +7325{\cdot}1 & -19015{\cdot}9 & -9445{\cdot}9 & +22008{\cdot}1 \end{Vmatrix}.$$

The matrix of sample correlation coefficients is therefore

$$\begin{Vmatrix} 1 & -0{\cdot}4788 & -0{\cdot}7885 & +0{\cdot}3617 \\ -0{\cdot}4788 & 1 & +0{\cdot}4807 & -0{\cdot}8685 \\ -0{\cdot}7885 & +0{\cdot}4807 & 1 & -0{\cdot}3759 \\ +0{\cdot}3617 & -0{\cdot}8685 & -0{\cdot}3759 & 1 \end{Vmatrix}.$$

Hence the multiple correlation coefficient of y_1 on y_2, y_3, y_4 is given by $R^2 = 0{\cdot}637$. Calculated by the approximation indicated in the last paragraph of the preceding section* $R^2 = 0{\cdot}656$; the true value obtained from equation (27) with variances in the ratio 1 : 3 : 1 : 4 is 0·727. The upper 10 and 5 % levels of significance, obtained from Thompson's tables with $\nu_1 = n-k = 6$, $\nu_2 = k-1 = 3$, are respectively 0·622 and 0·704. We find $L_1 = 0{\cdot}565$, the 5 and 1 % levels obtained from Nayer's (1936) tables being respectively 0·797 and 0·719, so that this test gives a more significant result than the one based on R^2. The relative merits of L_1 and the test we have provided, which cannot be judged on the results of one example, remain a problem to be investigated.

* I.e. calculated from 1 − (harmonic mean of g_{ii})/(arithmetic mean of g_{ii}).

SUMMARY

An exact test has been put forward for the equality of variances and covariances in any number k of 1- or 2-variate normal populations; the test is also exact for two 3- or 4-variate populations; but is restricted in application to equal sample sizes n from the k populations where n exceeds pk, p being the number of variates. The moments of the criterion are available for k p-variate populations where the statistic used is equivalent to that employed by Wilks (1935) to test the independence of two groups of variates (of sizes p and $p(k-1)$), and has the same distribution. In the univariate case the power of the test is known as a function of one parameter. Comparison with the L_1 criterion has already been made when $p=1$ and $k=2$, the tests being practically the same, and an example worked out of the use of the test when $p=1$.

Our thanks are due to E. C. Fieller for drawing our attention to the papers by Morgan and Pitman and suggesting that the test given there for the equality of two variances might be extended to more than two; also to Prof. E. S. Pearson for pointing out the need of certain explanatory additions.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

AITKEN, A. C. (1939). *Determinants and Matrices*. Oliver and Boyd, Edinburgh.

BISHOP, D. J. (1939). On a comprehensive test of the homogeneity of variances and covariances in multivariate problems. *Biometrika*, **31**, 31.

FISHER, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc.* A, **121**, 654.

GEARY, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, **27**, 310.

MORGAN, W. A. (1939). A test for the significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika*, **31**, 13.

NAIR, U. S. (1939). The application of the moment function in the study of distribution laws in statistics. *Biometrika*, **30**, 274.

NAYER, P. P. N. (1936). An investigation into the application of Neyman and Pearson's L_1 test, with tables of percentage limits. *Statist. Res. Mem.* **1**, 38.

NEYMAN, J. & PEARSON, E. S. (1931). On the problem of k samples. *Bull. int. Acad. Cracovie*, Série A, p. 460.

PEARSON, E. S. & NEYMAN, J. (1930). On the problem of two samples. *Bull. int. Acad. Cracovie*, Série A, p. 73.

PITMAN, E. J. G. (1939). A note on normal correlation. *Biometrika*, **31**, 9.

THOMPSON, C. M. (1941). Tables of percentage points of the incomplete beta-function. *Biometrika*, **32**, 168.

WALD, A. & BROOKNER, R. J. (1941). On the distribution of Wilks' statistic for testing the independence of several groups of variates. *Ann. Math. Statist.* **12**, 137.

WILKS, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471.

WILKS, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrika*, **3**, 309.

WILKS, S. S. (1943). *Mathematical Statistics*. Princeton University Press.

APPENDIX

As an illustration of the algebraic form of $W(k, 1)$ the Editor has suggested to me that it might be helpful to show the relation of the general formula (17) to the matrix G in this simple case when $k = 2$. Here, using a common notation for a sample mean

$$g_{11} = \sum_{\alpha=1}^{n} (x^1_{1\alpha} - x^1_{1\cdot})^2, \quad g_{22} = \sum_{\alpha=1}^{n} (x^2_{1\alpha} - x^2_{1\cdot})^2, \quad g_{12} = \sum_{\alpha=1}^{n} (x^1_{1\alpha} - x^1_{1\cdot})(x^2_{1\alpha} - x^2_{1\cdot}),$$

$$G = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix}, \quad \tilde{G} = \begin{bmatrix} g_{22} & -g_{12} \\ -g_{12} & g_{11} \end{bmatrix} \Big/ \{g_{11}g_{22} - g_{12}^2\},$$

$$S(2,1) = g_{11} + g_{22} + 2g_{12}, \quad \tilde{S}(2,1) = \frac{g_{11} + g_{22} - 2g_{12}}{g_{11}g_{22} - (g_{12})^2}.$$

Whence, using (17), (32) is at once obtained for $W(2, 1)$. For $k > 2$ the full expression for $\tilde{S}(k, 1)$ in terms of the g's is complicated and the matrix notation becomes essential.

THE ESTIMATION FROM INDIVIDUAL RECORDS OF THE RELATIONSHIP BETWEEN DOSE AND QUANTAL RESPONSE

By D. J. FINNEY

Lecturer in the Design and Analysis of Scientific Experiment, University of Oxford

1. Introduction

A type of biometric problem frequently encountered by the statistician is that which requires the estimation and study of a relationship between dose and response. 'Dose' is here a general term indicating the magnitude of a stimulus applied to certain test subjects, and 'response' is a measure of the effect which the stimulus produces on the subjects. When the test subjects are living matter, whether plants, animals or bacteria, pieces of tissue or single cells, the response to a specified dose is unlikely to be constant in repeated trials, and regression methods must be used in the estimation of the relationship.

In some classes of data, the response is 'all-or-nothing' or quantal, and cannot be measured quantitatively. Ordinary regression methods are then no longer applicable; methods based on the transformation of the proportion of subjects showing the response at any dose level to the normal equivalent deviate (Gaddum, 1933), or to the probit (Bliss, 1934*a*, *b*), however, have proved very powerful for simplifying the statistical analysis. In recent years, full accounts of the underlying theory of these transformations, and of their application, have been published by various authors (see, for example, Bliss, 1935*a*, *b*; Finney, 1947, 1948). An additional difficulty sometimes found is that the intensity of the stimulus cannot be *selected* in advance of a test, but can only be *measured* after the test has taken place; only rarely will two or more subjects happen to receive exactly the same dose, and more usually the records consist of a list of doses with, for each, a statement of whether a single subject receiving that dose responded or not.* For example, in some methods for the testing of insecticidal potency, poison bait is offered to individual insects; the dose received by any insect cannot be specified in advance, and must instead be measured as the amount of poison ingested.

Data from experiments of this kind do not give empirical values for the proportion of subjects responding at each dose level, except in the trivial sense that every dose shows either zero or 100 % responding. Nevertheless, as Bliss (1938) has pointed out, the probit method can still be applied to estimation of the dose-response relationship. He has given a numerical example, though without showing full details of the working, but has admitted that assessment of the error of estimation presents some theoretical difficulties (Finney, 1947, § 43). An interesting example of experimental results requiring this type of analysis has recently been brought to the notice of the writer by Mr R. W. Gilliatt. These introduce an additional complication, since the dose is expressed in terms of two measurements, and a probit plane (Finney, 1943) or other bivariate regression function must therefore be estimated. An account of the analysis, with computational details, may help those who have encountered analogous problems in biological or other investigations.

* When response does not involve death or serious alteration of the test subject, one subject may be used many times; the example discussed in this paper is an instance. The form of the data will be the same, though the interpretation may require that tolerance variation between and within subjects be distinguished.

2. THE DATA

Research in human physiology has demonstrated that, under carefully controlled experimental conditions, a transient reflex vaso-constriction in the skin of the digits may follow a single deep breath (Bolton, Carmichael & Stürup, 1936). Gilliatt (1947) has found that the response depends in part on the volume of air taken in by the subject. Plethysmographic measurement of the volume changes in a finger was used to indicate the occurrence of a response, but assessment of the degree of vaso-constriction, in order to relate this to the inspiratory stimulus, was not practicable. Thus the records obtained for each test show only

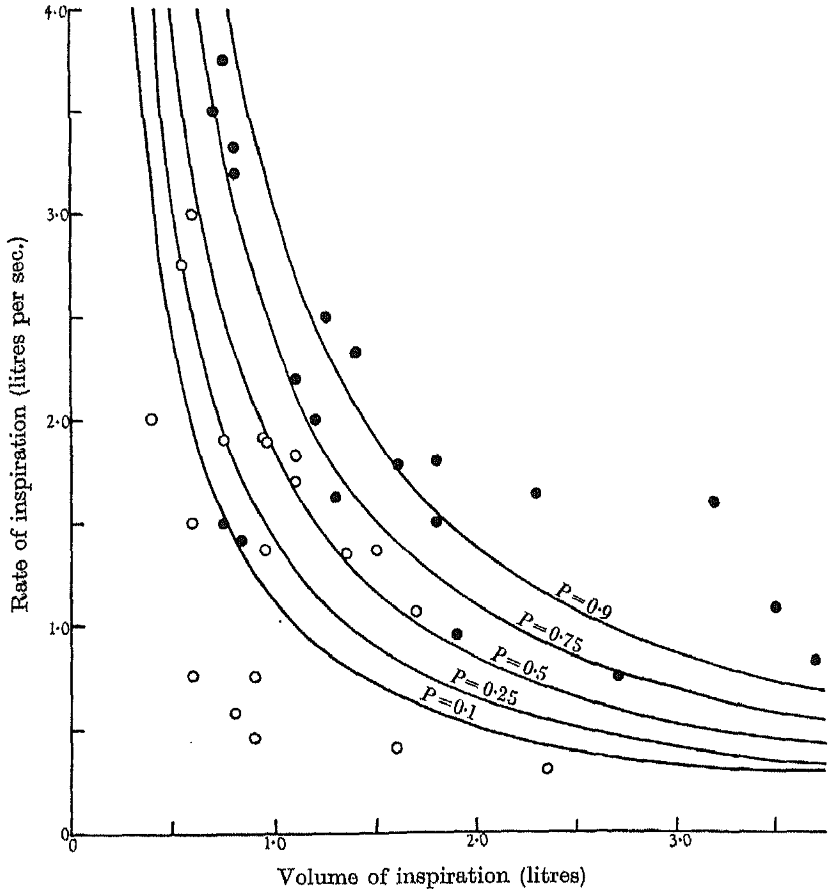


Fig. 1. Contours of dose-response surface for 0·1, 0·25, 0·5, 0·75 and 0·9 frequency of response, estimated from three-parameter equation. ○ no vaso-constriction; ● vaso-constriction.

the volume of air inspired, the average rate of inspiration, and whether or not vaso-constriction was produced. The above brief outline is sufficient for appreciation of the statistical problem, but a full account of the experimental procedure may be found in Gilliatt's paper; the results discussed here are presented in his Fig. 5.

The data, which Mr Gilliatt has kindly made available to the writer, were obtained from thirty-nine tests, in twenty of which vaso-constriction occurred. Tests were made on three different subjects, nine on D.W., eight on V.P.W., and twenty-two on S.J.S.; the results of the tests, with the subjects in this order, are shown in Table 1. In Fig. 1 are shown the thirty-nine combinations of volume in litres (V) and rate of inspiration in litres per second

Table 1. *Experimental data and details of calculations*

Volume in litres (V)	Rate in litres per sec. (R)	x_1	x_2	Response	Y	w	y	wx_1	wx_2	wy	P
3·7	0·825	1·57	0·92	+	6·8	0·18	7·26	0·2826	0·1656	1·3068	0·962
3·5	1·09	1·54	1·04	+	7·3	0·08	7·68	0·1232	0·0832	0·6144	0·989
1·25	2·5	1·10	1·40	+	6·5	0·27	7·02	0·2970	0·3780	1·8954	0·931
0·75	1·5	0·88	1·18	+	3·7	0·34	8·97	0·2992	0·4012	3·0498	0·098
0·8	3·2	0·90	1·51	+	5·8	0·50	6·53	0·4500	0·7550	3·2650	0·789
0·7	3·5	0·85	1·54	+	5·6	0·56	6·42	0·4760	0·8624	3·5952	0·741
0·6	0·75	0·78	0·88	−	1·3	0·00	1·05	0·0000	0·0000	0·0000	0·000
1·1	1·7	1·04	1·23	−	5·1	0·63	3·74	0·6552	0·7749	2·3562	0·530
0·9	0·75	0·95	0·88	−	2·4	0·04	2·06	0·0380	0·0352	0·0824	0·005
0·9	0·45	0·95	0·65	−	1·0	0·00	0·76	0·0000	0·0000	0·0000	0·000
0·8	0·57	0·90	0·76	−	1·4	0·00	1·14	0·0000	0·0000	0·0000	0·000
0·55	2·75	0·74	1·44	−	4·3	0·53	3·53	0·3922	0·7632	1·8709	0·248
0·6	3·0	0·78	1·48	−	4·8	0·63	3·72	0·4914	0·9324	2·3436	0·430
1·4	2·33	1·15	1·37	+	6·6	0·24	7·09	0·2760	0·3288	1·7016	0·950
0·75	3·75	0·88	1·57	+	6·0	0·44	6·66	0·3872	0·6908	2·9304	0·848
2·3	1·64	1·36	1·21	+	7·1	0·11	7·51	0·1496	0·1331	0·8261	0·982
3·2	1·6	1·51	1·20	+	8·0	0·01	8·30	0·0151	0·0120	0·0830	0·999
0·85	1·415	0·93	1·15	+	3·9	0·40	7·87	0·3720	0·4600	3·1480	0·128
1·7	1·06	1·23	1·03	−	5·2	0·63	3·72	0·7749	0·6489	2·3436	0·563
1·8	1·8	1·26	1·26	+	6·7	0·21	7·17	0·2646	0·2646	1·5057	0·958
0·4	2·0	0·60	1·30	−	2·6	0·06	2·23	0·0360	0·0780	0·1338	0·007
0·95	1·36	0·98	1·13	−	4·1	0·47	3·41	0·4606	0·5311	1·6027	0·179
1·35	1·35	1·13	1·13	−	5·1	0·63	3·74	0·7119	0·7119	2·3562	0·533
1·5	1·36	1·18	1·13	−	5·4	0·60	3·62	0·7080	0·6780	2·1720	0·662
1·6	1·78	1·20	1·25	+	6·3	0·34	6·86	0·4080	0·4250	2·3324	0·897
0·6	1·5	0·78	1·18	−	3·0	0·13	2·58	0·1014	0·1534	0·3354	0·025
1·8	1·5	1·26	1·18	+	6·2	0·37	6·79	0·4662	0·4366	2·5123	0·894
0·95	1·9	0·98	1·28	−	5·0	0·64	3·75	0·6272	0·8192	2·4000	0·489
1·9	0·95	1·28	0·98	+	5·2	0·63	6·28	0·8064	0·6174	3·9564	0·577
1·6	0·4	1·20	0·60	−	2·4	0·04	2·06	0·0480	0·0240	0·0824	0·005
2·7	0·75	1·43	0·88	+	5·6	0·56	6·42	0·8008	0·4928	3·5952	0·727
2·35	0·03	1·37	0·48	−	2·8	0·09	2·41	0·1233	0·0432	0·2169	0·015
1·1	1·83	1·04	1·26	−	5·2	0·63	3·72	0·6552	0·7938	2·3436	0·600
1·1	2·2	1·04	1·34	+	5·7	0·53	6·47	0·5512	0·7102	3·4291	0·767
1·2	2·0	1·08	1·30	+	5·8	0·50	6·53	0·5400	0·6500	3·2650	0·776
0·8	3·33	0·90	1·52	+	5·8	0·50	6·53	0·4500	0·7600	3·2650	0·806
0·95	1·9	0·98	1·28	−	5·0	0·64	3·75	0·6272	0·8192	2·4000	0·489
0·75	1·9	0·88	1·28	−	4·3	0·53	3·53	0·4664	0·6784	1·8709	0·243
1·3	1·625	1·11	1·21	+	5·4	0·60	6·34	0·6660	0·7260	3·8040	0·665
					Totals	14·29		14·9980	17·8375	74·9914	
					Means			1·0495	1·2483	5·2478	

Swx_1^2	Swx_1x_2	Swx_2^2	Swx_1y	Swx_2y	Swy^2
16·235606	18·338532	22·783383	79·738989	94·125032	433·586326
15·741078	18·721261	22·265669	78·706859	93·608054	393·541643
0·494528	−0·382729	0·517714	1·032130	0·516978	40·044683

(R), together with indications of whether or not the subject responded under these conditions. Inspection of Fig. 1 shows that, in general, when both V and R were small no response occurred, when either was large (unless the other was very small) the response occurred, and in an intermediate region the proportion of responses increased as either V or R increased. There was no sharply defined threshold separating combinations of V and R giving the response from those giving no response; instead, there appeared to be a probability of response ranging from practical certainty under some conditions to zero under others.

As an aid to fuller understanding of the influence of breathing on vaso-constriction, examination of the relationship between V, R, and the probability of response seemed desirable. Since so few observations were available for each subject, the data were unlikely to be sufficient to show differences between subjects; this point is discussed later, but in the main analysis the distinction between subjects is ignored. For any form of response assessment, the testing of one subject many times must introduce a danger that the result of one test will be affected not only by its own stimulus but by preceding stimuli and by the effects they produced. In this investigation, each subject was given a number of preliminary tests until he appeared to have settled into the routine. The observations recorded in Table 1 were obtained after these preliminary trials; they are tabulated in the order of testing, and show no indication of effects of previous history, but clearly such effects would have to be very pronounced if they were to be detectable on this amount of data.

3. METHOD OF ANALYSIS

Preliminary examination of the data suggested that the occurrence of a response was largely determined by the magnitude of VR, the product of volume and rate, curves on which the probability of response has a constant value being approximately hyperbolae of the form

$$VR = \text{constant}. \tag{1}$$

A little consideration shows that an equation of this type is more reasonable than an equation linear in V and R, though the data are almost certainly inadequate for discriminating between many alternative types of relationship that might be postulated. A system of curves similar to, but rather more general than, equation (1), namely,

$$V^{\beta_1}R^{\beta_2} = \text{constant}, \tag{2}$$

was selected for trial; this equation may alternatively be regarded as representing a series of parallel linear relationships

$$\beta_1 \log V + \beta_2 \log R = \text{constant} \tag{2a}$$

between the logarithms of volume and rate for a fixed probability of response.

A specified combination of V and R will not necessarily always give the same result (response or no response) with a subject, for, even though the subject is unaltered, minor uncontrolled variations in his environment may affect his susceptibility to the applied stimulus. For a particular value of V, the threshold value of R (the value which under the conditions prevailing at any instant would be just sufficient to produce a response) will have a frequency distribution; similarly, for a particular R, there will be a frequency distribution of threshold values of V. If these distributions may be taken as normal in $\log V$ and $\log R$, and, for simplicity, they are supposed to be such that the mean of either logarithm is linearly related to the selected value of the other, then the probability of response will be determined by an expression of the form

$$\beta_1 \log V + \beta_2 \log R,$$

and the threshold values of this quantity will be normally distributed. If x_1 and x_2 are written for $\log(10V)$ and $\log(10R)$ respectively (the factor of 10 is introduced in order to make x_1 and x_2 always positive), this statement enables the probability of response, P, to be expressed as

$$P = \int_{-\infty}^{\alpha+\beta_1 x_1+\beta_2 x_2} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2} du, \tag{3}$$

where α, β_1 and β_2 are parameters to be estimated from the data. The estimation may be regarded as the fitting of a probit regression plane, for Y, the probit of P being given by

$$Y = 5+\alpha+\beta_1 x_1+\beta_2 x_2. \tag{4}$$

Substitution of the value of Y corresponding to a specified probability gives the required linear relationship, equation (2*a*), between x_1 and x_2 for that probability, from which the estimated curves of constant probability, equation (2), may easily be derived.

The procedure for fitting a probit plane has been described elsewhere (Finney, 1943, 1947, § 31), and its chief features need no alteration for application to individual records. Providing that a first approximation to the equation can be guessed, repeated cycles of computation will give values for the parameters which approach more and more closely to the maximum likelihood estimates. Care in the choice of the first approximation will reduce the number of cycles needed; a poor choice will delay the convergence, though it will not affect the ultimate result. Since only a single observation is available for each combination of x_1 and x_2, every working probit is either a maximum or minimum value, according to whether or not the response occurs. When there is only one dose factor, in the fitting of a probit regression line to records of individuals, grouping of doses and treatment of the observations in a group as if they related to an average dose may reduce the labour of the early computing cycles, but, since it will tend to give an underestimate of the regression coefficient, the final cycle may need to use the detailed records. Bliss (1938) has given an example illustrating grouping of this kind. Grouping is less easily applied, however, when two or more dose factors have to be used, and, for the data under discussion, the individual records were used throughout except in the formation of the first approximation.

In the standard form of probit analysis, with moderately large numbers of observations at each level of dose, a χ^2 is usually computed for testing the significance of discrepancies between the data and the fitted equation; this χ^2 is numerically the same as would be obtained by calculation from expected and observed numbers of responses and non-responses for each dose. If there are few observations in any dose group, the expected number of responses or of non-responses (or of both) is likely to be small, and, as is well known, χ^2 may then fail to follow the sampling distribution tabulated for that statistic. Data of the type under discussion here are extreme examples of this situation, the number of observations for each dose being reduced to unity, so that any disturbance of the χ^2 distribution is likely to be encountered in its most acute form. No complete theoretical investigation of this matter has yet been made, but the practical implications are discussed more fully in § 5.

On the assumption that the estimate of equation (4) is an adequate representation of the data, lines of constant response probability may be obtained for any specified probability; these may be plotted according to equation (2) on a V, R scale. Standard statistical processes also enable fiducial limits to be assigned to the position of any of these curves. The difficulty of dealing with the estimation of error for individual records, and the inadequacy of the data for any sensitive test of whether equation (2) is a satisfactory representation of the

system of curves, throw doubts on the exact interpretation of these fiducial limits. Nevertheless, they give some idea of the confidence that can be attached to the estimated curves, at least for moderate values of V and R; for extremes of either measurement, far more extensive data would be needed before much faith could be placed in the fitted equation.

4. COMPUTATIONS FOR ESTIMATING THE THREE-PARAMETER EQUATION

In this and the two succeeding sections, the computations for Gilliatt's data will be described in detail. The first five columns of Table 1 show the thirty-nine pairs of values of V and R which occurred in the experiments, followed by the corresponding values of x_1 and x_2, together with a statement of whether or not the subject responded. Before the probit computations could be initiated, a first approximation to equation (4) was needed; this was obtained with the aid of the suggestion, from the plotting of the data shown in Fig. 1, that the constant probability curves were approximately the hyperbolae of equation (1), or alternatively

$$x_1 + x_2 = \text{constant}.$$

As Bliss (1938) has pointed out, there is no objection to the use of overlapping groups in the formation of the first approximation. The data were therefore grouped according to the value of (x_1+x_2), as shown below, and the proportion of responses in each group was obtained from Table 1:

x_1+x_2	Responses	Proportion (p)	Probit of p	First approximation
1·5–1·9	0/7	0·00	—	3·3
1·6–2·0	0/7	0·00	—	3·6
1·7–2·1	2/7	0·29	4·4	3·9
1·8–2·2	2/9	0·22	4·2	4·2
1·9–2·3	3/14	0·21	4·2	4·5
2·0–2·4	8/19	0·42	4·8	4·8
2·1–2·5	13/24	0·54	5·1	5·1
2·2–2·6	17/25	0·68	5·5	5·4
2·3–2·7	16/17	0·94	6·6	5·7
2·4–2·8	12/12	1·00	—	6·0

Each proportion was regarded as an estimate for the median value of (x_1+x_2) in the group, i.e. 1·7, 1·8, 1·9, ..., and its probit was read from one of the standard tables (Finney, 1947, Table I; Fisher & Yates, 1947, Table IX). As may be seen above, these probits were fairly well fitted by the guessed equation

$$Y = -1{\cdot}8 + 3(x_1+x_2), \tag{5}$$

which was therefore used as a first approximation to equation (4).

A first set of expected probits was calculated from equation (5), and inserted as Y in an earlier version of Table 1. A cycle of routine probit calculations, just as described in the next two paragraphs, then led to an improved approximation to the required estimate, on which a second cycle of improvement was based. The figures shown in Table 1 relate to the fourth of these cycles, based upon the approximation

$$Y = -9{\cdot}127 + 6{\cdot}666x_1 + 5{\cdot}906x_2 \tag{6}$$

from the third cycle. Equation (6) is very different from equation (5), suggesting that more care might have been given to the selection of a first approximation; that the grouping

adopted would lead to underestimation of the regression coefficients was expected, but insufficient allowance for this was made. Of course the 'improvement' in the approximations refers to their approach to the solution of the maximum likelihood equations, and is not necessarily always an approach to the true relationship.

The column of expected probits, Y, in Table 1 was calculated by substitution of pairs of values x_1, x_2 in equation (6); one decimal place here is quite sufficient. The weighting coefficient, w, for each observation was then read from tables (Finney, 1947, Table II; Fisher & Yates, 1947, Table XI) and entered in its column. The working probit, y, takes a maximum value for every observation giving a response and a minimum value for every observation giving no response, since these give empirical rates of 100 % and zero respectively; values of y were read directly from Finney's table (1947, Table III; or, less simply for the minimum values, from Fisher & Yates, 1947, Table XI). The numbers of decimal places shown for the entries in Table 1 are sufficient for data of this type; indeed possibly one decimal for w and for y would be enough. Columns wx_1, wx_2, and wy were then filled, and the weighted sums of squares and products of deviations, required for the calculation of the regression of y on x_1 and x_2, were completed at the bottom of the table.

The equations giving the estimates of the regression coefficients, b_1 and b_2, are

$$0{\cdot}494528b_1 - 0{\cdot}382729b_2 = 1{\cdot}032130,$$
$$-0{\cdot}382729b_1 + 0{\cdot}517714b_2 = 0{\cdot}516978.$$

Later calculations use the variances and covariance of b_1 and b_2; the equations were therefore solved by first obtaining the matrix inverse to that formed by the coefficients of b_1 and b_2 (Finney, 1943, 1947, § 31; Fisher, 1946, § 29). This matrix is

$$V = \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} = \begin{pmatrix} 4{\cdot}726144 & 3{\cdot}493883 \\ 3{\cdot}493883 & 4{\cdot}514482 \end{pmatrix}; \tag{7}$$

the accuracy of the data is insufficient to need the number of decimal figures shown here, but their retention assists the checking and maintains the internal consistency of the analysis. Now

$$b_1 = 1{\cdot}032130v_{11} + 0{\cdot}516978v_{12}$$
$$= 6{\cdot}68426,$$

and similarly

$$b_2 = 5{\cdot}94003.$$

The estimate of equation (4) is then

$$Y = \bar{y} + b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2)$$

or

$$Y = -9{\cdot}182 + 6{\cdot}6843x_1 + 5{\cdot}9400x_2, \tag{8}$$

a result which differs little from equation (6) and may be regarded as a sufficiently close approximation to the maximum likelihood estimate. Since

$$b_2/b_1 = 0{\cdot}889, \tag{9}$$

equation (8) may be transformed to give

$$VR^{0{\cdot}889} = \text{constant} \tag{10}$$

as the relationship estimated to exist between V and R for a specified probability; the value of the constant can be obtained by substitution of the probit of the probability in equation (8), a process which gives 1·10, 1·36, 1·71, 2·16 and 2·66 for probabilities of 10, 25, 50, 75 and 90 % respectively. Typical contours have been drawn in Fig. 1 so as to indicate the form of the relationship.

5. GOODNESS OF FIT

When probit analysis is applied to data containing many observations in each dose group, the weighted sum of squares of deviations between the empirical probits and the predictions from the fitted equations is a χ^2, with degrees of freedom equal to the number of dose groups reduced by the number of fitted parameters. If S_{uv} is written for the weighted sum of products of deviations of variates u and v, application of this method here would give

$$\begin{aligned}\chi^2_{[36]} &= S_{yy} - b_1 S_{x_1 y} - b_2 S_{x_2 y} \\ &= 40\cdot045 - 6\cdot6843 \times 1\cdot0321 - 5\cdot9400 \times 0\cdot5170 \\ &= 30\cdot08. \end{aligned} \tag{11}$$

When the dose groups are small, however, the χ^2 so calculated cannot be trusted as an indicator of the significance of deviations from the fitted equation, and it is presumably most unreliable when each group is reduced to a single observation. Apart from slight discrepancies caused by imperfect approximation to the maximum likelihood solution, the χ^2 in equation (11) is algebraically identical with that which would be derived, by the usual form of calculations, from comparison of observed numbers responding and not responding in each group with expectations computed from the fitted equation. As is well known from the study of contingency tables, when the expectations in some classes are small the sampling distribution of such a χ^2 may be very different from that shown in the standard tables (Finney, 1947, Table VI; Fisher & Yates, 1947, Table IV); with data from individual records, no class can have an expectation greater than unity, and for many the expectation will be very much less, so that the discrepancy from the tabulated χ^2 distribution is likely to be serious.

The general effect of small expectations on the random sampling distribution of χ^2 appears to be that the mean value remains about equal to the number of degrees of freedom, but that the variance in repeated sampling is increased. Consequently, samples from a population according with the null hypothesis are likely to show an excess of very high and very low values, as judged by the tables of χ^2. Thus there is little danger that significant evidence of deviations from expectation will be overlooked in an uncritical application of the test, though apparently significant values of χ^2 need to be examined with care before they are regarded as evidence sufficient to justify rejection of the null hypothesis. Low values, as in Gilliatt's data 30 with 36 degrees of freedom, need cause little alarm, for they clearly indicate no serious deviation from expectation. High values may in the first instance be compared with the standard tables of the χ^2 distribution; if they fall beyond the significance level, a closer examination should be made before judging the null hypothesis to be untenable, for the apparent significance may be due to large contributions from one or two aberrant points. Gilliatt's data provide an illustration of this. The expected probits for each pair of values of x_1 and x_2 in Table 1 have been calculated from equation (8), and the probabilities, P $(= 1 - Q)$, corresponding to these have been entered in the last column of the table; P is then the expectation of the number of responses for each dose. The χ^2 obtained from the observed and expected numbers in seventy-eight classes is easily seen to be the sum of Q/P for all doses giving a response, plus P/Q for all giving no response. Inspection of the column for P shows small contributions to χ^2 everywhere, except for two instances of responses with probabilities of only 0·098 and 0·128, contributing 9 and 7 respectively; clearly the occurrence of these two responses as the most extreme events in thirty-nine trials need not be regarded

as serious evidence against the null hypothesis. The result of calculating χ^2 by this more laborious process is a total of 30·3, which agrees closely with that already given in equation (11).

One method of modifying a χ^2 test so as to remove its extreme sensitivity to deviations from small expectations is to combine expected and observed frequencies over several adjacent groups, so as to obtain groups with larger expectations; the number of degrees of freedom is then taken as the number of remaining groups less the number of fitted parameters. Of course the groups must be chosen objectively, and without regard to the agreement between the frequencies. The statistic still will not follow the χ^2 distribution exactly, but the approximation should be fairly satisfactory under the usual restriction that the groups be so chosen that none of the expected frequencies is small. This procedure often has to be adopted in probit analysis because of small expectations at very low or very high doses (Finney, 1947, § 18). With individual records, however, only very extensive grouping will give expectations sufficiently large for the χ^2 test to be trusted; the reduction of a large χ^2 to a value below the significance level might then appear indicative of an insensitive test rather than of absence of serious discrepancies.*

Probably no completely satisfactory solution of the difficulty is to be expected. Individual records usually arise from experimental work in which the obtaining of large numbers of observations presents considerable difficulty. Often the whole series will consist of less than fifty observations, and, unless previous information enables the range of doses to be chosen satisfactorily, many of the observations will be made at doses for which response is either almost certain or almost impossible. Even if the individual dose-tolerances could be measured directly, a test of normality of their distribution (which is what the χ^2 test attempts to provide) could not be very sensitive when based on only fifty measurements; if, instead, only quantal data are available, indicating merely whether a dose is below or above the tolerance value, a sensitive normality test is still less likely to exist (Finney, 1947, § 43).

Gilliatt's data, a series of only thirty-nine observations, provide an extreme instance of the difficulty of formulating a sensitive test of goodness of fit. Nevertheless, an attempt has been made to examine the discrepancies between the observations and the null-hypothesis expressed by equation (4). In Table 2 are compared the observed and expected frequencies when the data are grouped according to the value of $VR^{0\cdot 889}$. This is equivalent to a grouping based on the value of Y, the expected probit in equation (8), and, as this quantity had been evaluated for each observation in order to give P, it was used in the construction of Table 2. Since three parameters have been estimated from the data, four groups is the least number for giving a χ^2 test. The limits of the groups were chosen so as to give similar numbers of observations in each. Inspection of Table 2 shows that the groups are still too small for a χ^2 test to be trusted, thus suggesting that the data are inadequate for any useful test of goodness of fit to be made. The only anomaly in Table 2 is the occurrence of two responses where the expectation is 0·3, and this is clearly insufficient to cause much worry.

The inadequacy of the data for detecting any differences in sensitivity between the three subjects may be seen from Table 3. The first nine entries in Table 1 relate to D.W., and sum-

* In his discussion of the analysis of individual records, Bliss (1938) suggests adjustment of the χ^2 test, not by altering the calculation of the statistic but by reducing the number of degrees of freedom allotted to it; he gives an empirical rule for the reduction, based upon the expectations in terminal dose groups. This method, however, not only lacks any theoretical basis, but seems liable to have an effect opposite to that which is needed; it will attribute significance to high values of χ^2 even more readily than will the unadjusted test.

mation of the values of P gives the expected number of responses for this subject; similarly the next eight and the last twenty-two entries give the numbers for V.P.W. and S.J.S. respectively. Inspection of Table 1 shows that the tests on each subject were fairly widely distributed over the range of values of x_1 and x_2. Table 3 shows excellent agreement between totals of observed and expected responses for each subject, thus suggesting that any individual differences that exist are small by comparison with the variation in sensitivity of the same subject in different tests.

Table 2. *Comparison of observed and expected frequencies of response*

Range of Y	Frequencies of results				
	Observed −	Observed +	Total	Expected −	Expected +
–4	8	2	10	9·72	0·28
4–5	6	0	6	3·92	2·08
5–6	5	8	13	4·26	8·74
6–	0	10	10	0·59	9·41
Total	19	20	39	18·49	20·51

Table 3. *Comparison of subjects*

Subject	Frequencies of results				
	Observed −	Observed +	Total	Expected −	Expected +
D.W.	3	6	9	4·0	5·0
V.P.W.	4	4	8	3·5	4·5
S.J.S.	12	10	22	11·0	11·0
Total	19	20	39	18·5	20·5

6. LIMITS OF ERROR

The variances of b_1 and b_2 and the covariance between them are respectively v_{11}, v_{22} and v_{12} as defined in equation (7). Hence the variance of Y, the expected probit corresponding to any pair of values x_1, x_2, is

$$V(Y) = \frac{1}{Sw} + v_{11}(x_1 - \bar{x}_1)^2 + 2v_{12}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) + v_{22}(x_2 - \bar{x}_2)^2, \tag{12}$$

where Sw is the sum of the w column in Table 1. All these variances are derived from binomial probability distributions. In the usual form of probit analysis, with a batch of subjects at each dose, the precision of the estimated relationship between dose and response is discussed as though the variation were normal, an assumption which is justifiable on account of the large numbers of individuals involved. Here, with only thirty-nine observations in all, the

assumption is less safe, but may be adopted for lack of any more trustworthy method of dealing with the data. It is unlikely to be seriously misleading, except possibly for extreme levels of the response probability, P.

Equation (12) may now be used in the assignment of fiducial limits to any one of the curves of equal probability given by equation (10). For suppose that t is the normal deviate corresponding to the significance level to be used in defining the fiducial limits, and that Y_0 is the probit of a probability P_0. Then for any values of x_1, x_2 for which

$$(Y - Y_0)^2 > t^2 V(Y),$$

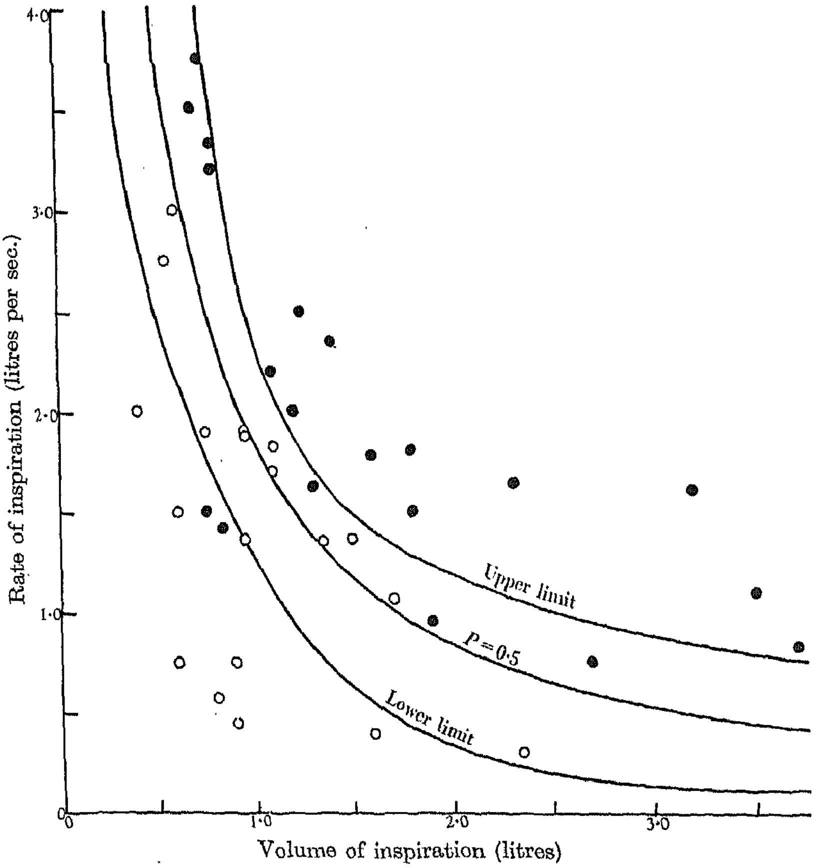


Fig. 2. Fiducial limits (5 % probability) to 0·5 frequency contour of Fig. 1.
O no vaso-constriction; ● vaso-constriction.

where Y is determined from equation (8), the expected probit differs significantly from Y_0, and for values of x_1, x_2 which reverse the inequality the difference is not significant. Therefore the equation

$$(Y - Y_0)^2 = t^2 V(Y) \tag{13}$$

gives the limiting values of (x_1, x_2) for which the null hypothesis that the true expected probit is Y_0 is not untenable in the light of the data; in other words, equation (13) defines curves in the (x_1, x_2) plane which are fiducial limits to the estimated locus of points having a constant response probability P_0. These curves are clearly hyperbolae. In Figs. 2 and 3, the 5 % fiducial limit curves ($t = 1{\cdot}960$) for $P_0 = 0{\cdot}5$ and $P_0 = 0{\cdot}9$ respectively have been plotted in

the (V, R) plane; details of the calculation need not be given here, but Fig. 2, for example, is derived from the equation

$$(14{\cdot}182 - 6{\cdot}6843x_1 - 5{\cdot}9400x_2)^2 = 3{\cdot}841\left[\frac{1}{14{\cdot}29} + 4{\cdot}7261(x_1 - 1{\cdot}0495)^2 + 6{\cdot}9878(x_1 - 1{\cdot}0495)(x_2 - 1{\cdot}2483) + 4{\cdot}5145(x_2 - 1{\cdot}2483)^2\right].$$

The pairs of curves are like hyperbolae in form. That for $P_0 = 0{\cdot}5$ defines a band on either side of the estimated relationship which is quite narrow for moderate values of V and R

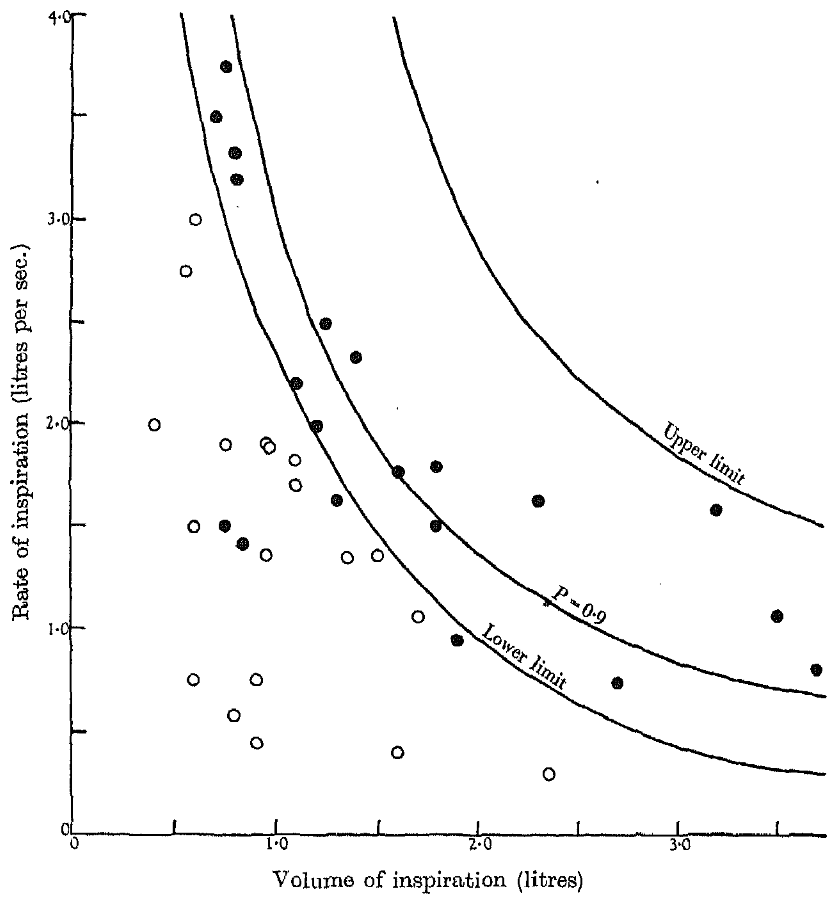


Fig. 3. Fiducial limits (5 % probability) to 0·9 frequency contour of Fig. 1.
○ no vaso-constriction; ● vaso-constriction.

though naturally it widens considerably at the extremes. That for $P_0 = 0{\cdot}9$, as might be expected from general consideration of the problem, allows much greater uncertainty on the side of high values of V and R; similarly, for $P_0 = 0{\cdot}1$, that band would be relatively wider on the side of low values of V or R.

The curves shown in Figs. 1, 2 and 3 may be regarded as plane sections, for selected values of Y, of a three-dimensional diagram relating Y to V and R. In terms of x_1, x_2 instead of V, R, this diagram is the three-dimensional analogue of the familiar diagram showing a regression line with hyperbolic curves indicating limits of error on either side; the line

generalizes to a plane, and the limits are now defined by two sheets of a hyperboloid, one above and one below the plane.

The theoretical basis of the curves illustrated in Figs. 2 and 3 is perhaps insecure, but undoubtedly they give a useful indication of the dependence of the probability of response on V and R and of the reliability of the estimation of this relationship. Much as an experimenter might wish for a more precise assessment of the effects of V and R, experience suggests that results such as those obtained here are as good as can be expected from a total of thirty-nine quantal observations.

7. The two-parameter equation

In § 3, the equation

$$VR = \text{constant} \tag{1}$$

was suggested as an expression of the curves of constant response probability, but the more complex equation (2) was adopted for use in §§ 4–6. There are no theoretical reasons for believing that equation (1) represents the true form of the relationship, and the more general form was chosen in order that the complete calculations might be illustrated. The values of b_1 and b_2 obtained, however, do not differ very greatly by comparison with the standard error of their difference; in fact

$$V(b_1 - b_2) = v_{11} - 2v_{12} + v_{22}$$
$$= 2{\cdot}253,$$

and therefore

$$b_1 - b_2 = 0{\cdot}744 \pm 1{\cdot}501.$$

In the absence of any significant difference between the regression coefficients, the common scientific procedure of preferring the simpler hypothesis (Occam's Razor) suggests that equation (4) might be replaced by

$$Y = \alpha + \beta(x_1 + x_2). \tag{14}$$

For the estimation of equation (14), the computations are similar to, but shorter than, those of § 4, since $(x_1 + x_2)$ may be replaced by a single variate, x, and a simple regression calculated; the calculations in § 4 were used to give a first set of expected probits, from which was derived the estimate

$$Y = -9{\cdot}475 + 6{\cdot}4067(x_1 + x_2). \tag{15}$$

Only two parameters have been estimated from the data, and calculation as for equation (11) gives

$$\chi^2_{[37]} = 28{\cdot}76.$$

The difference between the two χ^2 values may be taken as a further criterion of whether or not the extra parameter is needed, closely related to the test of significance of $(b_1 - b_2)$;

$$\chi^2_{[1]} = 1{\cdot}32$$

is not significant, though again the validity of the χ^2 test is in doubt.

Substitution of the probit of a specified probability in equation (15) gives the value of the constant in equation (1). For the 50 % response probability, for example, the constant is 1·82; over the range of values tested, the curves

$$VR^{0{\cdot}889} = 1{\cdot}71 \quad \text{and} \quad VR = 1{\cdot}82$$

differ only slightly. Similarly, fiducial limits to $(x_1 + x_2)$ may be calculated, for any Y_0, as upper and lower values of the product VR. No special interest attaches to these calculations; the novelties due to the individual records are exactly as for the three-parameter equation discussed in earlier sections, and otherwise the method is entirely that of ordinary probit analysis (Finney, 1947, Chapter 4). For comparison with the three-parameter equation, diagrams similar to Figs. 2 and 3 may be prepared; both the constant probability curves and

the fiducial limits are then true hyperbolae. Fig. 4 shows the results for a 50 % response probability, and is to be compared with Fig. 2. The constant probability curve in Fig. 4 differs little from that in Fig. 2, though naturally the difference increases for large values of V or R where the curves are less well determined. For moderate values of V and R, the fiducial

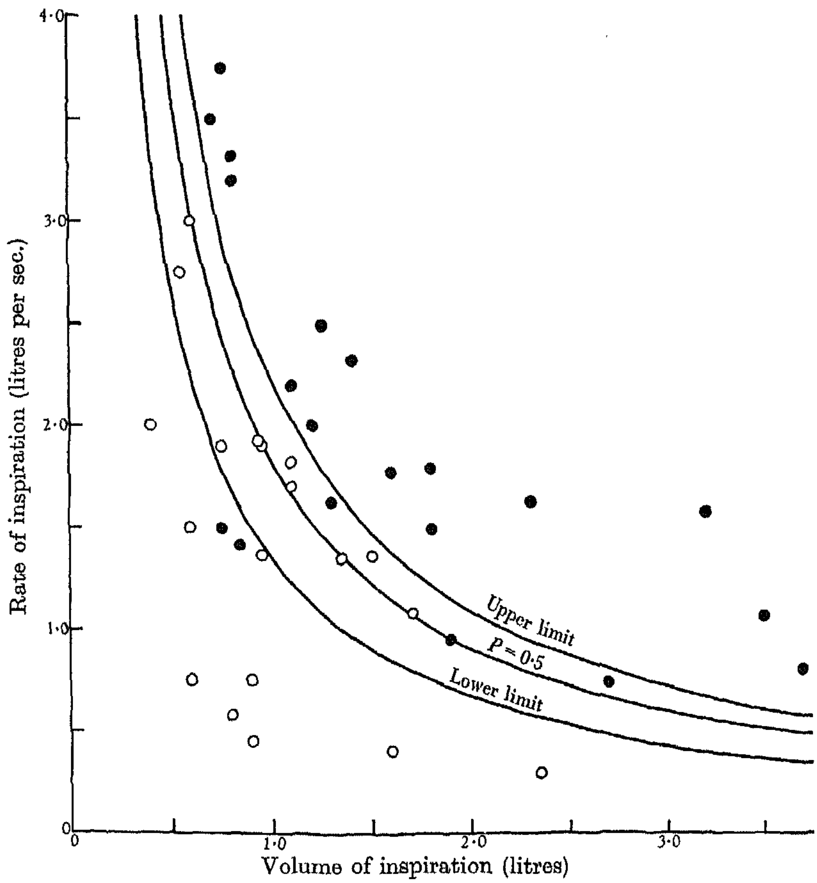


Fig. 4. Contour of dose-response surface for 0·5 frequency of response, estimated from two-parameter equation, and its 5 % fiducial limits (compare Fig. 2). ○ no vaso-constriction; ● vaso-constriction.

limit curves are practically the same as the corresponding curves in Fig. 2, but for more extreme values they lie much closer to the curve of constant probability; since the data show no significant difference between b_1 and b_2, it is to be expected that a more precisely estimated relationship between stimulus and response will be obtained if an assumption that $\beta_1 = \beta_2$ is made, so that the information on the two regression coefficients can be combined, and this shows itself by narrowing the zone of error for the constant probability curve.

8. SUMMARY

The method of probit analysis has been developed to assist the study of the relationship between the magnitude of a stimulus and the proportion of tests in which a particular quantal response to that stimulus appears. In some research problems, the stimulus cannot be controlled sufficiently to make possible the administration of a specified magnitude, though the stimulus actually received by any one subject can later be measured. It will then seldom happen that two subjects receive exactly the same 'dose', and the data for

statistical analysis will generally consist of a series of doses with, for each, a statement of whether or not a single subject showed the characteristic response.

Even for data of this type, the probit transformation can aid the estimation of the relationship between dose and the probability of response. The calculations leading to the estimate are more tedious than is usual in probit analysis, because of slow convergence from a provisional equation to the final form, but follow the usual pattern. The validity of the χ^2 test of goodness of fit (in reality a test for the normality of distribution of individual tolerances) must be doubted, however, since the disturbance due to small class numbers will be encountered in its most extreme form. Extensive grouping of results for adjacent doses will provide a test less open to objection, though this will generally be insensitive to all but the grossest deviations from normality; indeed, no valid sensitive test is to be expected with individual records unless these are very numerous.

In this paper, the calculations have been illustrated on data relating to a reflex vaso-constriction which sometimes occurs in the skin of the digits of human subjects after a single deep breath. The relationship between the occurrence of this response and two dose factors, the volume and the rate of inspiration, has been estimated for the combined records from three subjects; inclusion of two dose factors complicates the analysis, since a bi-variate regression equation must be fitted, but does not affect the underlying theory. The χ^2 test has been discussed at length, though there is no indication of non-normality or of heterogeneity of the data. The reliability with which the dependence of the probability of response on the dose factors is estimated has also been examined, and curves bounding fiducial regions, within which the true probability contours may confidently be asserted to lie, have been determined. This method of representing the limits of error is applicable to other forms of probit analysis involving two dose factors and is not restricted to individual records, though it has not previously been described.

I am indebted to Mr R. W. Gilliatt, of the Department of Physiology, both for permission to make use of his data in an illustration of the statistical methods of my paper and for assistance in describing his experimental procedure. My thanks are due also to Miss M. Callow, who prepared Figs. 1–4.

REFERENCES

Bliss, C. I. (1934*a*). The method of probits. *Science*, **79**, 38–9.

Bliss, C. I. (1934*b*). The method of probits—a correction. *Science*, **79**, 409–10.

Bliss, C. I. (1935*a*). The calculation of the dosage-mortality curve. *Ann. Appl. Biol.* **22**, 134–67.

Bliss, C. I. (1935*b*). The comparison of dosage-mortality data. *Ann. Appl. Biol.* **22**, 307–33.

Bliss, C. I. (1938). The determination of dosage-mortality curves from small numbers. *Quart. J. Pharm.* **11**, 192–216.

Bolton, B., Carmichael, E. A. & Stürup, G. (1936). Vaso-constriction following deep inspiration. *J. Physiol.* **86**, 83–94.

Finney, D. J. (1943). The statistical treatment of toxicological data relating to more than one dosage factor. *Ann. Appl. Biol.* **30**, 71–9.

Finney, D. J. (1947). *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve.* Cambridge: University Press.

Finney, D. J. (1948). The principles of biological assay. *J.R. statist. Soc. Suppl.* **9**, 46–91.

Fisher, R. A. (1946). *Statistical Methods for Research Workers*, 10th ed. Edinburgh: Oliver and Boyd.

Fisher, R. A. & Yates, F. (1947). *Statistical Tables for Biological, Agricultural and Medical Research*, 3rd ed. Edinburgh: Oliver and Boyd.

Gaddum, J. H. (1933). Reports on biological standards. III. Methods of biological assay depending on a quantal response. *Spec. Rep. Ser. Med. Res. Coun., Lond.*, no. 183.

Gilliatt, R. W. (1947). Vaso-constriction in the finger following deep inspiration. *J. Physiol.* (in the Press).

A POWER FUNCTION FOR TESTS OF RANDOMNESS IN A SEQUENCE OF ALTERNATIVES

By F. N. DAVID

1. During recent years attention has been focused on what might be called the 'group' test for randomness in a sequence of alternatives. Thus, if E denote the happening of an event, and $\bar{E}$ its negation, the number of alternations of E and $\bar{E}$ in a sequence supposedly random has been chosen as a test criterion. This test has been put to different uses by W. L. Stevens (1939), A. Wald & J. Wolfowitz (1940) and F. N. David (1947). It seems worth while therefore to enquire what is the power of this test against a set of specifically defined alternate hypotheses. The hypothesis to be tested will be that there is randomness within the sequence, with the alternate hypothesis that if there is no randomness then there is dependence of the type found in a simple Markoff chain. The same procedure will hold good for dependence of the types found in more complex chains although in these cases the enumeration is a little troublesome.

2. If there is a sequence of dependent events

$$E_1, E_2, E_3, \ldots, E_n,$$

then it is an elementary proposition of the probability calculus that

$$P\{E_1E_2E_3\ldots E_{n-1}E_n\} = P\{E_1\}\,P\{E_2 \mid E_1\}\,P\{E_3 \mid E_1E_2\}\ldots P\{E_n \mid E_1E_2\ldots E_{n-1}\}.$$

If the events are independent, then

$$P\{E_1E_2E_3\ldots E_{n-1}E_n\} = P\{E_1\}\,P\{E_2\}\,P\{E_3\}\ldots P\{E_n\}.$$

This relation will be the basis of H_0, the hypothesis to be tested. If there is dependence as in a simple Markoff chain, then mathematically each event will be dependent on the event immediately preceding it, but will be independent of any of the other events. In this case we shall have

$$P\{E_1E_2E_3\ldots E_{n-1}E_n\} = P\{E_1\}\,P\{E_2 \mid E_1\}\,P\{E_3 \mid E_2\}\ldots P\{E_n \mid E_{n-1}\}.$$

This relation will be the basis of H_1, the hypothesis alternate to H_0.

3. For the hypothesis, H_0, let the probability that an event E will occur in a single trial be p, and let the probability of $\bar{E}$ (the negation of E), be q, where $p+q=1$. The probability of obtaining any given sequence of r_1 E's and r_2 $\bar{E}$'s will be

$$p^{r_1}q^{r_2}.$$

The number of ways in which r_1 E's and r_2 $\bar{E}$'s may be arranged to form $2t$ and $2t+1$ sets of E's and $\bar{E}$'s alternately is

$$f_{2t} = \frac{2(r_1-1)!\,(r_2-1)!}{(t-1)!\,(t-1)!\,(r_1-t)!\,(r_2-t)!} \quad \text{and} \quad f_{2t+1} = f_{2t} \times \frac{r_1+r_2-2t}{2t}.$$

Writing $k=2t$ or $2t+1$ as desired, the probability of obtaining a sequence of r_1 E's and r_2 $\bar{E}$'s arranged in k sets is

$$P\{k \mid r_1, r_2, H_0\} = \frac{p^{r_1}q^{r_2}f_k}{\sum\limits_{\text{All }t} p^{r_1}q^{r_2}f_t} = \frac{f_k}{\sum\limits_{\text{All }t} f_t}.$$

k may take values $2, 3, \ldots, 2r_2$, if $r_1 = r_2$, and values $2, 3, \ldots, 2r_2+1$ if $r_1 > r_2$.

4. Following the orthodox procedure, in order to test the hypothesis, H_0, it is necessary to find two numbers k_1 and k_2 such that

$$P\{k<k_1 \mid H_0\} \leqslant \tfrac{1}{2}\epsilon, \quad P\{k>k_2 \mid H_0\} \leqslant \tfrac{1}{2}\epsilon,$$

and therefore

$$P\{k_1 \leqslant k \leqslant k_2\} \geqslant 1-\epsilon,$$

where ϵ is a number arbitrarily at choice. If an observed number of sets, say k', falls outside the limits k_1 and k_2 then the hypothesis H_0 will be rejected in favour of some alternate hypothesis, H_1. Alternately if H_0 is not true, but H_1 is, then

$$1-P\{k_1 \leqslant k \leqslant k_2 \mid H_1\}$$

will be the power of the test in the sense of the word as used by Neyman & Pearson. Whether k_1 or k_2 is chosen to judge the significance of an observed k' will depend on which departure from randomness it is most important not to overlook. If the alternate hypothesis is that there is positive dependence in the chain, i.e. that E having occurred in the sth trial it is more likely to occur in the $(s+1)$st trial, then k_1 would be chosen. Such a situation was envisaged in a proposed smooth test to supplement the χ^2 criterion (David, 1947). If, however, the alternate hypothesis is that there is negative dependence, i.e. that E having occurred in the sth trial, it is less likely to occur in the $(s+1)$st trial, then k_2 would be the appropriate criterion. If it is immaterial whether the departure from randomness is positive or negative dependence, then both k_1 and k_2 may be used.

5. We now consider the alternate hypothesis, H_1. Write E_s for the occurrence of the event E in the sth trial and $\bar{E}_s$ for its negation. Let

$$P\{E_1\} = P, \qquad P\{\bar{E}_1\} = Q, \quad P+Q = 1 \text{ and } P \geqslant Q,$$
$$P\{E_s \mid E_{s-1}\} = p_1, \quad P\{\bar{E}_s \mid E_{s-1}\} = q_1,$$
$$P\{E_s \mid \bar{E}_{s-1}\} = p_2, \quad P\{\bar{E}_s \mid \bar{E}_{s-1}\} = q_2.$$

Thus p_1 and q_2 are probabilities of no change and p_2 and q_1 probabilities of a change. If the events are independent then

$$p_1 = p_2 = P \quad \text{and} \quad q_1 = q_2 = Q.$$

6. In calculating the probability of obtaining any given sequence, what will matter will be the number of changes from E to $\bar{E}$ and back again. Let $f_t(r_1)$ be the number of ways in which r_1 E's can be arranged in t groups, i.e. let

$$f_t(r_1) = \frac{(r_1-1)!}{(t-1)!\,(r_1-t)!}.$$

If there are $2t$ groups in a sequence of r_1 E's and r_2 $\bar{E}$'s, the number of ways of obtaining such a sequence will be

$$f_t(r_1)\, f_t(r_2)$$

if the sequence starts with E or with $\bar{E}$. The probability of obtaining any given sequence of r_1 E's and r_2 $\bar{E}$'s of $2t$ groups will be

$$Pp_2^{t-1}p_1^{r_1-t}q_1^t q_2^{r_2-t} \quad \text{or} \quad Qq_1^{t-1}q_2^{r_2-t}p_2^t p_1^{r_1-t}.$$

This follows from the fact that a sequence of $2t$ groups beginning with E will imply t changes from E to $\bar{E}$ and $t-1$ changes from $\bar{E}$ to E. The changes are reversed in number if the sequence starts with $\bar{E}$. For $2t+1$ groups the number of ways of obtaining the sequence will be

$$f_{t+1}(r_1)\, f_t(r_2) \quad \text{or} \quad f_t(r_1)\, f_{t+1}(r_2)$$

according as the sequence begins with E or $\bar{E}$. The respective probabilities will be

$$Pp_2^t p_1^{r_1-t-1}q_1^t q_2^{r_2-t} \quad \text{and} \quad Qq_1^t q_2^{r_2-t-1}p_2^t p_1^{r_1-t}.$$

The probability therefore of obtaining a sequence of $r_1 E$'s and $r_2 \bar{E}$'s in $2t$ groups will be therefore, under hypothesis H_1,

$$P\{2t \mid r_1 r_2 H_1\} = \frac{\left(\frac{p_2 q_1}{p_1 q_2}\right)^t f_t(r_1) f_t(r_2) \left(\frac{P}{p_2} + \frac{Q}{q_1}\right)}{\sum_{t=1}^{r_2} \left(\frac{p_2 q_1}{p_1 q_2}\right)^t \left[f_t(r_1) f_t(r_2) \left(\frac{P}{p_2} + \frac{Q}{q_1}\right) + \frac{P}{p_1} f_{t+1}(r_1) f_t(r_2) + \frac{Q}{q_2} f_t(r_1) f_{t+1}(r_2) \right]}.$$

The probability of obtaining $r_1 E$'s and $r_2 \bar{E}$'s in $2t+1$ groups will be similarly

$$P\{2t+1 \mid r_1 r_2 H_1\} = \frac{\left(\frac{p_2 q_1}{p_1 q_2}\right)^t \left[\frac{P}{p_1} f_{t+1}(r_1) f_t(r_2) + \frac{Q}{q_2} f_t(r_1) f_{t+1}(r_2) \right]}{\sum_{t=1}^{r_2} \left(\frac{p_2 q_1}{p_1 q_2}\right)^t \left[f_t(r_1) f_t(r_2) \left(\frac{P}{p_2} + \frac{Q}{q_1}\right) + \frac{P}{p_1} f_{t+1}(r_1) f_t(r_2) + \frac{Q}{q_2} f_t(r_1) f_{t+1}(r_2) \right]}.$$

7. So far no mention has been made of any possible connexion between p_1, q_1, p_2 and q_2. It is obvious in all cases we shall have

$$p_1 + q_1 = 1, \quad p_2 + q_2 = 1,$$

but the connexion between p_1 and p_2 is not immediate. We shall make the simplifying assumption which is perhaps most closely related to practical problems, and shall state that where nothing is known about the $s-1$ trials preceding the sth trial, $P\{E_s\} = P$ and $P\{\bar{E}_s\} = Q$. Under this assumption we have

$$p_2 = \frac{Pq_1}{Q}, \quad q_2 = 1 - \frac{Pq_1}{Q}.$$

This result is reached easily by noticing that

$$P\{E_s\} = P\{E_s E_{s-1}\} + P\{E_s \bar{E}_{s-1}\} = P\{E_{s-1}\} P\{E_s \mid E_{s-1}\} + P\{\bar{E}_{s-1}\} P\{E_s \mid \bar{E}_{s-1}\}$$

whence
$$P = Pp_1 + Qp_2.$$

8. The alternative hypothesis chosen to illustrate the power function formulae is that there is positive dependence in the sequence, i.e. k_1 is found so that

$$P\{k \leqslant k_1 \mid H_0\} \leqslant \epsilon \quad \text{and} \quad 1 - P\{k \geqslant k_1 \mid H_1\}$$

is calculated, when $p_1 \geqslant P$. For economy of drawing, several power curves or what are really sections of a kind of power surface, plotted to coordinates P, p_1, have been put together in the diagrams of Fig. 1. For example the bottom left-hand diagram shows for $r_1 = r_2 = 10$ sections of the conditional power surface for $P = 0{\cdot}5$, $0{\cdot}6$ and $0{\cdot}75$. When H_0 is true and $P = p_1$, we have the 5 % risk of rejecting H_0 wrongly. As $p_1 - P$ increases the chance of detecting the fact increases, but in a way dependent on P. The other three diagrams show similar sections of the surfaces with $r_1 = r_2 = 5$, with $r_1 = 14$, $r_2 = 6$ and with $r_1 = 7$, $r_2 = 3$. In practice it will not be known what the value of P is, but the curves show reasonably well how the power of the test varies as P and p_1 (and therefore p_2) vary. It is clear that the test for randomness under discussion is most powerful when the numbers of alternates are equal, i.e. when $r_1 = r_2$. The power declines sharply when r_1 increases at the expense of r_2. Another point which emerges is that the test is only moderately powerful, against the given alternate hypothesis tested, when $r_1 + r_2 = 20$, and it would appear therefore that if it was desired not to overlook a possible departure from randomness in the form of positive dependence in the chain, then the length of the sequence should consist of at least 20 units. The question of other possible tests we shall not discuss at this stage.

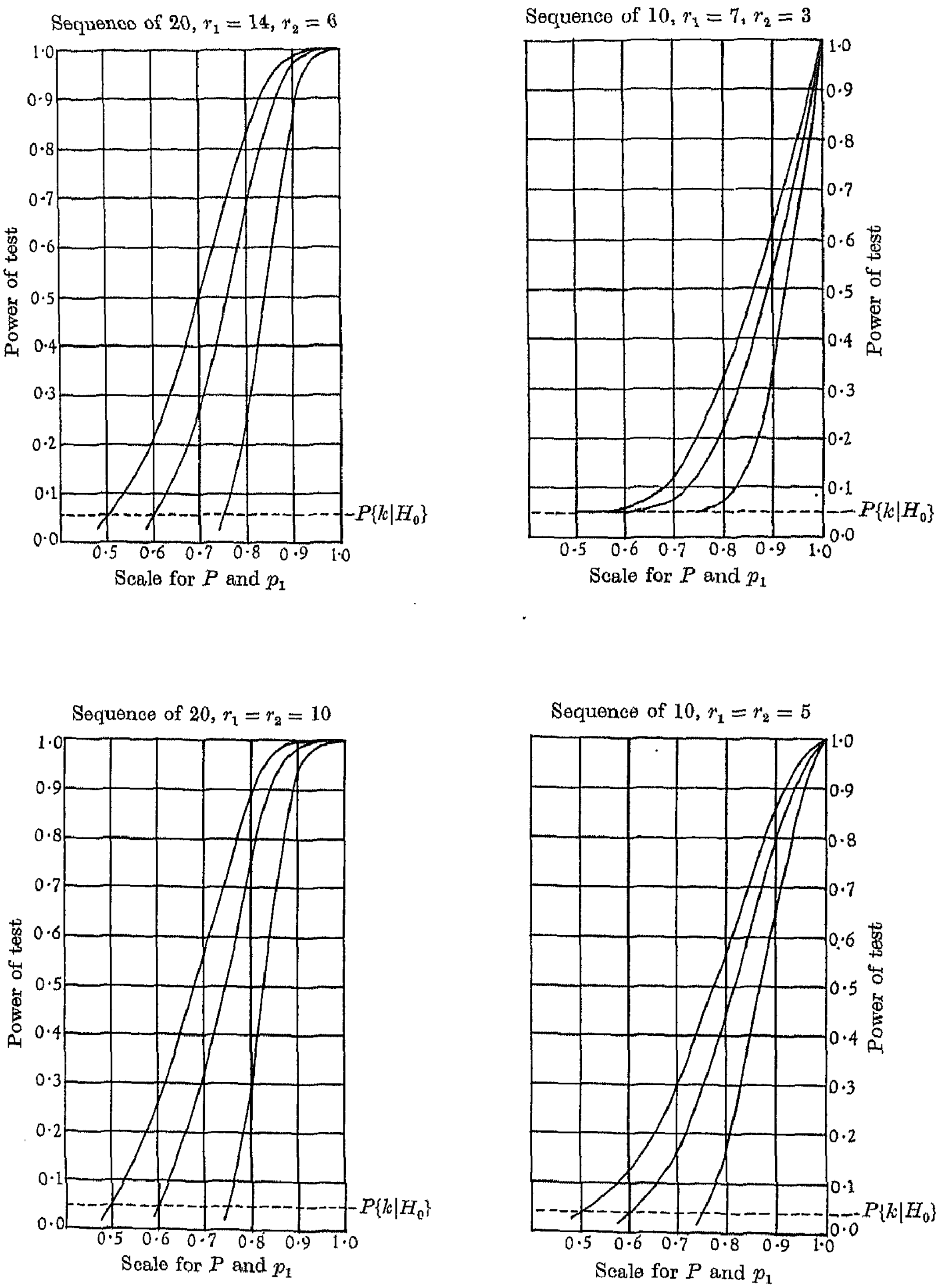


Fig. 1. Conditional power curves when the alternate hypothesis is positive dependence.

9. It will be noticed that $P\{2t \text{ or } 2t+1 \mid r_1 r_2 H_1\}$ which have been loosely termed power function formulae are not power functions in the sense originally defined by Neyman & Pearson, but they appear to involve a justifiable extension of that idea. In order to distinguish them from the usual meaning of the words power function, I shall refer to them as *conditional power functions*. The theory of the conditional power function may be stated briefly in the following way. It is assumed that all possible samples (or sequences) may be classified according to their composition. Suppose that there are k of these mutually exclusive classes, which are also the only possible, say $C_1, C_2, \ldots, C_k$. We have considered only the case where k is finite but it appears likely that the method can be extended to cover the case where k is enumerably infinite. These classes, $C_1, C_2, \ldots, C_k$ will correspond to regions forming a partition of the sample space.

Let H_0 be the hypothesis tested and w_0 be the critical region used for the rejection of this hypothesis. Given that a sample is in C_i (say), and that an alternate hypothesis H_1 is true, then the probability that H_0 will be rejected is

$$P\{E\epsilon w_0 C_i \mid E\epsilon C_i, H_1\} = \frac{P\{E\epsilon w_0 C_i \mid H_1\}}{P\{E\epsilon C_i \mid H_1\}}$$

where $w_0 C_i$ means the region common to w_0 and to C_i and, following the Neyman-Pearson notation, E is the sample point. Regarded as a function of H_1 this is the conditional power function of the test associated with w_0 in the subset C_i of samples.

The Neyman-Pearson power function, which we might call here the *overall* power function, will be

$$P\{E\epsilon w_0 \mid H_1\} = \sum_{i=1}^{k} P\{E\epsilon w_0 C_i \mid H_1\} = \sum_{i=1}^{k} P\{E\epsilon w_0 C_i \mid E\epsilon C_i, H_1\} P\{E\epsilon C_i \mid H_1\},$$

which may be looked on as a weighted average of the conditional power functions.

10. There seems to be no reason why w_0 should not be built up of portions $w_0 C_i$, these portions being chosen to maximize each term of the summation, i.e. $w_0 C_i$ chosen to maximize the conditional power function. For example, to revert to the specific case of randomness within a sequence with which we have been dealing, the different partitions of r $(= r_1 + r_2)$ are the mutually exclusive and only possible classes C_i. It is conceivable, although practically not very likely, that for each of these classes there will exist a different test which is more powerful to detect specifically defined departures from the basic hypothesis tested than any other test. The decision as to which is the most powerful test, against the same specifically defined alternatives, to use for any given class will be decided by the conditional power function. Once this has been decided the procedure for the complete test of significance may be laid down. This will be: (i) count the number of alternatives in the sequence, i.e. find r_1 and r_2, (ii) from (i) decide the appropriate test of significance to use, (iii) apply the test. The power of the test as laid down by (i), (ii) and (iii), in the usual meaning of the word, will be given by the overall power function.

It is proposed to discuss these, and other applications of the conditional power function technique, in a further publication. I have been concerned here with trying to explain what I believe to be the basic ideas, and to forestall possible criticism that I am falling into error (of the third kind) and am choosing the test falsely to suit the significance of the sample.

REFERENCES

STEVENS, W. L. (1939). *Ann. Eugen., Lond.*, **9**, 11.
WALD, A. & WOLFOWITZ, J. (1940). *Ann. Math. Statist.* **11**, 147.
DAVID, F. N. (1947). *Biometrika*, **34**, 299.

A NUMERICAL SOLUTION OF THE PROBLEM OF MOMENTS

By H. O. HARTLEY and S. H. KHAMIS

1. Introduction

Given a statistical variable x and its frequency distribution $f(x)$, then, under certain continuity conditions for $f(x)$, the moments

$$\mu_r = \int x^r f(x)\,dx \quad (r = 0, 1, 2, \ldots) \tag{1}$$

can be evaluated for any integer r. For certain distributions $f(x)$ the integrations in (1) can be carried out analytically resulting in simple formulae for the moments. In general there is no inherent difficulty in obtaining numerical values for the moments by numerical quadrature.

The inverse problem is to find the distribution $f(x)$ given the moments μ_r. This problem, commonly known as 'The Problem of Moments', has received considerable attention by mathematicians and is of interest in statistical distribution theory. There are numerous statistics for which it is difficult to obtain a formula of the random sampling distribution $f(x)$ amenable to numerical evaluation. On the other hand, in such cases it is often possible to find simple formulae for the random sampling moments (Bartlett, 1937). Sometimes such formulae are available for *all* integer r; more often than not, however, μ_r is only known for a limited number of small r (e.g. $r = 0, 1, \ldots, 6$). A simple method of 'determining' $f(x)$ from the given moments would therefore be helpful in such cases.

Examples of variables of this kind are the numerous moment statistics or k-statistics for which random sampling moments can be evaluated, notably by R. A. Fisher's (1929, 1930) combinatorial methods, whilst their exact sampling distributions are usually unknown. As related statistics we should mention here the moment ratios $\sqrt{b_1}$ and b_2 used in tests for deviation from normality (Geary, 1947, Geary & Worlledge, 1947). For these, the low-order moments are known exactly. A similar situation arises with statistics defined as likelihood ratios, as, for instance, with the criterion L_1 required for testing heterogeneity in a set of variances. Moments for this statistic were obtained by Neyman & Pearson as early as 1931, yet, although approximations to $f(L_1)$ have been obtained (Bartlett, 1937; Hartley, 1940; Nayer, 1936; Neyman & Pearson, 1931; Sukhatme, 1936; Welch, 1935, 1936), there is still considerable doubt about their accuracy in certain cases, and the exact formula obtained by Nair (1936) in the case of equal sample sizes is very complex.

These and numerous other problems of distribution point to the necessity of developing a numerical technique to deal with the following situation:

(i) A random variable x ranging between a and b (where a may be $-\infty$ and b may be $+\infty$) has a distribution function $f(x)$ known to have a continuous derivative of order n.

(ii) The moments

$$\mu_r = \int_a^b x^r f(x)\,dx \quad (r = 0, 1, \ldots, R), \tag{2}$$

are known numerically to any decimal accuracy desired but for a limited number of positive integers r, viz. $r = 0, 1, \ldots, R$. With the knowledge about $f(x)$ limited to the above conditions, is it possible to obtain numerical values for the probability integral $P(x) = \int_a^x f(x)\,dx$

depending on the moments only, and is it possible to make a statement on the accuracy of these values in terms of the derivatives of the function $f(x)$?

Problems of this kind have hitherto been treated principally in two ways:

(a) When $R = 2, 3$ or 4 nothing better can be expected than a 'good fit', which is often achieved by fitting the appropriate Pearson-type curve.

(b) With R in the neighbourhood of 5–8, expansions of the Gram Charlier, Laguerre or Jacobi type have been used, either as cumulant or as moment expansions. Such theorems as are available for statements on the convergence and asymptotic behaviour of these expansions usually require too many moments to be known. Often the expansions are only asymptotic, and unless the distribution is close to the generating curve (Normal for Gram Charlier, Γ for Laguerre), the results are often disappointing (see, for example, Kendall, 1945, Chapter 6).

2. OUTLINE OF PRESENT METHOD

The method to be developed here is a direct application of finite-difference calculus and therefore provides both numerical answers to the problem, as well as gauges of their accuracy in form of remainder terms. The method is, in fact, closely linked with interpolation technique. When using any of the well-known interpolation formulae no mathematically rigorous statement on the accuracy of the interpolates can be made unless the magnitude of the remainder term can be estimated, and for this some knowledge about (say) the nth derivative of the function is required. Yet, in using such formulae the convergence of the difference table inspires confidence that 'the results of the interpolation can be accepted as a working hypothesis' (Milne Thomson, 1933, p. 62). Similarly, with the present method we shall give a numerical procedure of obtaining values of the probability integral. Certain checks of internal consistency will be described which inspire confidence that the answers are correct, but no rigorous statement on the accuracy can, of course, be made if this is to be based on a finite number of moments alone. The *exact* remainder terms which we derive will entail the high-order derivatives of $f(x)$, and it is hoped, in a second communication, to derive some *general* statements concerning their order of magnitude.

In order to simplify the argument we assume in this section that the range of x is finite (a and b finite).

The aim is to determine the probability integral of x, $P(x) = \int_a^x f(x)\,dx$ in tabular form, i.e. we wish to determine numerical values of

$$P_i = P(x_i) = \int_a^{x_i} f(x)\,dx \tag{3}$$

for discrete values of x_i. For convenience the group intervals $x_{i+1} - x_i$ will generally be chosen equidistant (group interval $= h$), and the number of intervals will be $R+1$, i.e. equal to the number of given moments (including $\mu_0 = 1$). Hence

$$x_i = a + ih, \quad h = (b-a)/(R+1). \tag{4}$$

The first differences in the table derived from equation (3) are the quantities

$$f_i = P_i - P_{i-1} = \int_{x_{i-1}}^{x_i} f(x)\,dx, \tag{5}$$

and are the familiar 'frequencies' f_i in a grouped frequency distribution with equidistant intervals (see Fig. 1). The link between these frequencies and the exact moments μ_r is then

established by the well-known formulae for Sheppard's correction. Using Kendall's (1938, 1945) derivation and remainder term, but extending his notation, we have

$$\sum_{i=1}^{R+1} f_i \xi_i^r = \mu_r + C(r,h) + S(r,h), \tag{6}$$

where the centre points ξ_i are given by

$$\xi_i = a + (i - \tfrac{1}{2})h \quad (i = 1, \ldots, R+1), \tag{7}$$

$C(r,h)$ denotes Sheppard's corrective term, viz.

$$C(r,h) = \sum_{j=1}^{[\frac{1}{2}r]} \left(\frac{h}{2}\right)^{2j} \binom{r}{2j} \frac{1}{2j+1} \mu_{r-2j}, \tag{8}$$

and $S(r,h)$ the remainder term.

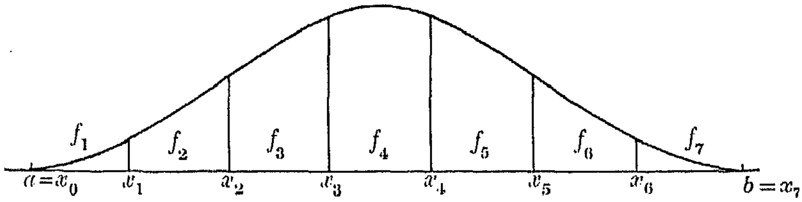


Fig. 1

The aim, now, is to use equations (6) to determine the unknown f_i from the given μ_r. To this end the remainder $S(r,h)$ must be examined: Most distribution functions have what is commonly known as high contact at the terminals of the variate range. This means that $f(x)$, as well as all its derivatives up to order, say, m, vanish at both ends of the range, i.e.

$$f^{(i)}(a) = f^{(i)}(b) = 0 \quad (i = 0, \ldots, m). \tag{9}$$

If for such functions we define $f(x) = 0$ outside the range $a \leqslant x \leqslant b$, it will have continuous derivatives of up to order m for $-\infty < x < +\infty$. It can then be shown that the remainder term is of the form (see, for example, Kendall, 1945, p. 69)

$$S_m(r,h) = -\frac{(R+1)h^m}{m!} B_m k^{(m)}(r,h,\theta_r) \quad (m \text{ even}), \tag{10}$$

$$S_m(r,h) = \frac{2(R+1)h^m}{m!} B_{m+1}^{(1)}(\tfrac{1}{2}) k^{(m)}(r,h,\theta_r) \quad (m \text{ odd}), \tag{11}$$

$$a \leqslant \theta_r \leqslant b,$$

where the B_j are the Bernoulli numbers, the $B_j^{(1)}$ are the Bernoulli polynomials of first order, the integrand function $k(r,h,x)$ is defined by

$$k(r,h,x) = x^r \int_{-\frac{1}{2}h}^{+\frac{1}{2}h} f(x+\xi)\, d\xi, \tag{12}$$

and its derivatives with regard to x are denoted by $k^{(i)}$. In the subsequent sections we shall assume (9) to hold (contact of order m), but will discuss the case when (9) is not satisfied in § 10.

The remainder term $S_m(r,h)$ will usually be small (see, for example, Kendall, 1945, p. 72). We shall therefore, in what follows, ignore $S_m(r,h)$ but will discuss the error thereby committed in § 5.

If, then, in (6) we omit $S(r,h)$ we obtain a system of $R+1$ linear equations for the $R+1$ unknowns f_i

$$\sum_{i=1}^{R+1} f_i \xi_i^r = \mu_r + C(r,h) = \bar{\mu}_r. \tag{13}$$

The matrix of this system of equations (v_R say) is of the form $|\xi_i^r|$ and has a classical determinant $\|\xi_i^r\|$, sometimes referred to as Vandermonde's determinant and well known to be $\neq 0$. The system can therefore be inverted once and for all and, for any *particular* case, the unknown f_i can then be determined by substituting the right-hand sides of (13), i.e. $\bar{\mu}_r$ in the inverse matrix v_R^{-1}. Denoting the elements of this inverse matrix by u_{ir} we have the system of equations

$$f_i = \sum_{r=0}^{R} u_{ir}\bar{\mu}_r. \tag{14}$$

Progressive addition of the f_i yields the P_j from $P_j = \sum_{i=1}^{j} f_i$* and therefore a table of $P(x)$ at interval h. Finally, intermediate values of $P(x)$ can be obtained by standard interpolation. Alternatively, as described in § 7, we may obtain directly a table of $P(x)$ at interval $\frac{1}{2}h$.

3. THE STANDARD FORM OF THE NUMERICAL INVERSION

The rank of the original matrix v_R is obviously equal to $R+1$, i.e. the number of moments given, whilst its elements are the powers of the centre points ξ_i^r. It is desirable therefore that, for any given R, scale and location of the variable x be transformed into a standard form X, so that only *one* matrix V_R and therefore only *one* matrix V_R^{-1} need be calculated for each R. It is most convenient to standardize as follows:

$$X = (x - \tfrac{1}{2}(a+b))\frac{R+1}{b-a} \quad (R \text{ even}), \tag{15}$$

$$X = (x - \tfrac{1}{2}(a+b))\frac{R+1}{b-a} + \frac{1}{2} \quad (R \text{ odd}). \tag{16}$$

It will be seen, therefore, that the range of X is $R+1$ and the group interval

$$H = X_{i+1} - X_i = 1.$$

From the given moments of x those of X (M_r say) about $X = 0$ can, of course, be calculated by the usual binomial formulae, and in what follows we assume that values of M_r are given numerically. Further, in analogy to (13), we have

$$\bar{M}_r = M_r + C(r, 1). \tag{17}$$

From (15) and (16) we obtain for the new centre points

$$\left.\begin{aligned} \Xi_i &= -\tfrac{1}{2}R, \ldots, 0, \ldots, +\tfrac{1}{2}R \quad \text{for even } R, \\ \Xi_i &= -\frac{R-1}{2}, \ldots, 0, \ldots, \frac{R+1}{2} \quad \text{for odd } R, \end{aligned}\right\} \tag{18}$$

and the matrix V_R becomes $|(i-1-\frac{1}{2}R)^r|$ or $|(i-\frac{1}{2}R-\frac{1}{2})^r|$. Thus, if the first six moments are given, we obtain for V_6:

$$V_6 \equiv \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ 9 & 4 & 1 & 0 & 1 & 4 & 9 \\ -27 & -8 & -1 & 0 & 1 & 8 & 27 \\ 81 & 16 & 1 & 0 & 1 & 16 & 81 \\ -243 & -32 & -1 & 0 & 1 & 32 & 243 \\ 729 & 64 & 1 & 0 & 1 & 64 & 729 \end{vmatrix}. \tag{19}$$

* It is, of course, possible to construct a matrix yielding the P_j directly from the $\bar{\mu}_r$, but we are here satisfied with determining the f_i first, as they are of independent interest.

In practice the important range of R will be from 5 to about 8. The inverse matrix V_6^{-1} is given below, and it is hoped to give V_7^{-1}, V_8^{-1} and V_5^{-1} in a subsequent paper. The inverse matrix V_6^{-1}, the elements of which are denoted by U_{ir}, can be written in the form

$$c_i f_i = \sum_{r=0}^{R} U'_{ir} \bar{M}_r, \tag{20}$$

where $U'_{ir} = c_i U_{ir}$, i.e. the c_i are suitable common denominators of the $U_{i0}, \ldots, U_{iR}$, and the U'_{ir} are given in the body of the schedule below:

		$\bar{M}_0 = 1$	$\bar{M}_1$	$\bar{M}_2$	$\bar{M}_3$	$\bar{M}_4$	$\bar{M}_5$	$\bar{M}_6$ = multiplier of column
i	$c_i f_i$	$r = 0$	1	2	3	4	5	6
1	$720 f_1$	0	-12	4	15	-5	-3	1
2	$120 f_2$	0	18	-9	-20	10	2	-1
3	$48 f_3$	0	-36	36	13	-13	-1	1
4	$36 f_4$	36	0	-49	0	14	0	-1
5	$48 f_5$	0	36	36	-13	-13	1	1
6	$120 f_6$	0	-18	-9	20	10	-2	-1
7	$720 f_7$	0	12	4	-15	-5	3	1

(21)

In order to use the above system of equations it would be necessary to compute the $\bar{M}_r$ from the given M_r, using formula (17). It is obviously more convenient to evaluate, once and for all, a matrix U''_{ir} giving the f_i directly in terms of the given M_r. This matrix is given below for $R = 6$:

i	f_i	$M_0 = 1$	M_1	M_2	M_3	M_4	M_5	M_6
1	f_1	0·000 379	$-$0·011 719	0·002 344	0·017 361	$-$0·005 208	$-$0·004 167	0·001 389
2	f_2	$-$0·005 227	0·109 375	$-$0·034 896	$-$0·152 778	0·072 917	0·016 667	$-$0·008 333
3	f_3	0·059 161	$-$0·683 594	0·618 490	0·253 472	$-$0·244 792	$-$0·020 833	0·020 833
4	f_4	0·891 373	0	$-$1·171 875	0	0·354 167	0	$-$0·027 778
5	f_5	0·059 161	0·683 594	0·618 490	$-$0·253 472	$-$0·244 792	0·020 833	0·020 833
6	f_6	$-$0·005 227	$-$0·109 375	$-$0·034 896	0·152 778	0·072 917	$-$0·016 667	$-$0·008 333
7	f_7	0·000 379	0·011 719	0·002 344	$-$0·017 361	$-$0·005 208	0·004 167	0·001 389

(22)

Working rule: Each f_i is obtained by forming the sum of seven products using the seven coefficients in the ith line and applying them to $M_0, \ldots, M_6$, e.g. $f_1 = 0{\cdot}000\,379 M_0 - 0{\cdot}011\,719 M_1 + \ldots + 0{\cdot}001\,389 M_6$.

4. Calculation of the incomplete B-function $I_x(8, 6)$ from its first six moments

As an example for the above method we consider the Beta Distribution for $p = 8$ and $q = 6$, viz.

$$f(x) = [B(8, 6)]^{-1} x^7 (1-x)^5.$$

Using the moments for this distribution about $x = 0$, $\mu_r = B(x+r, 6)/B(8, 6)$ $(r = 0, \ldots, 6)$ and transforming to the standard scale $X = 7x - 3{\cdot}5$, we obtain for the moments of X about $X = 0$: $M_1 = 0{\cdot}5$, $M_2 = 1{\cdot}05$, $M_3 = 1{\cdot}225$, $M_4 = 2{\cdot}77426$, $M_5 = 4{\cdot}41360$ and $M_6 = 10{\cdot}56942$. Substituting these in the matrix (22) we obtain values of f_i whose progressive sums are shown in Table 1 (calculated $I_x(8, 6)$). These may be compared with the 'exact' values obtained (by interpolation) from the *Tables of the Incomplete B-function* (1934). The worst discrepancy is about 2 in the fourth decimal. Higher accuracy can, of course, be obtained if the number of moments $(R+1)$ and therefore the number of f_i increases (see, for example, § 8, where the normal curve is obtained to 5-decimal accuracy).

A rather gratifying feature of the comparison is the higher *decimal* accuracy in the tails of the distribution. This is a consequence of the sensitivity of the higher moments to changes in the tail frequencies. Note also that the elements in the top and bottom lines of the inverse matrix (22) are much smaller than those in the other lines, so that any error in the right-hand sides of (13) has a smaller effect on the terminal f_i.

Table 1. *Comparison of 'calculated' and 'exact' values of* $I_x(8, 6)$

X	x	Exact I_x	Calculated I_x	Difference 10^{-5}
-2.5	1/7	0.000 11	0.000 09	2
-1.5	2/7	0.013 41	0.013 54	-13
-0.5	3/7	0.140 17	0.139 95	22
0.5	4/7	0.489 63	0.489 81	-18
1.5	5/7	0.862 61	0.862 70	-9
2.5	6/7	0.994 11	0.993 95	16
3.5	7/7	1.000 00	1.000 00	0

It might be argued that a further error will arise when determining intermediate values of I_x by interpolation in the 'calculated' table. This difficulty could, however, be overcome by shifting the grid of group intervals and using a standard X-scale with group end-points corresponding to the odd multiples of 1/14 in x, thereby obtaining I_x at points half-way between the arguments of Table 1. Such a method has actually been used in § 7.

5. The remainder term

A formal representation of the remainder term is immediately obtained by reverting to the exact equations (6). If we are concerned with distribution functions having contact of order m at the terminals, the error contributions to the f_i are obtained by substituting the $R+1$ remainder terms $S_m(r, h)$ ((10), (11)) in the inverse matrix v^{-1}. It is convenient to use the standard variate X-scale, $H = 1$ and the V^{-1} matrix when it will be found that

$$\text{error } f_i = \sum_{r=0}^{R} U_{ir} S_m(r, 1), \tag{23}$$

where $S_m(r, 1)$ is given by (10) or (11) putting $h = 1$ and remembering that the integrand function k must be taken in terms of the standard variate X, viz.

$$k(r, 1, X) = X^r \int_{-\frac{1}{2}}^{+\frac{1}{2}} f\left(\frac{b-a}{R+1}(X+\xi)\right) \frac{b-a}{R+1} d\xi. \tag{24}$$

Since the arguments θ_r of $k^{(m)}(r, 1, X)$ are unknown it will as a rule be necessary to substitute their respective maxima in (23), at the same time taking $|U_{ir}|$ in place of U_{ir}.

Although with (23) we have given a formal solution of the error term involved, in a manner similar to the remainder terms of interpolation formulae, it will in practice be difficult to estimate the magnitude of the error from this formula. It is hoped, therefore, to go into this aspect more fully in a second paper.

6. Infinite variate range and artificial truncation

When the range of the variate is infinite, i.e. when $a = -\infty$ and/or $b = +\infty$, it is, of course, possible to transform the variate x by, say, $y = y(x)$ such that the range of y is finite. However, in general, we shall not be able to assume that the moments of y are known or that they can

be derived from those of x. It is therefore necessary to adapt our method to deal with an infinite variate range. We shall treat here the case $b = +\infty$, the case $a = -\infty$ being identical and the case $a = -\infty$ *and* $b = +\infty$ being analogous.

For an infinite variate range, the condition of high contact is now replaced by

$$\lim_{x\to\infty} f^{(i)}(x) = 0 \quad (i = 0, 1, \ldots, m), \tag{25}$$

which results in remainder terms analogous to (10) and (11)*. Similarly, in equations (6) which correspond to Kendall's (1945) equations (3·40), the summation now extends from $i = 1$ to $i = \infty$, there being an infinity of frequencies $f_i = \int_{x_{i-1}}^{x_i} f(x)\,dx$. Now since the μ_r exist we know that

$$\int_a^\infty x^r f(x)\,dx \tag{26}$$

is convergent. Accordingly

$$\lim_{b\to\infty} \sum_{i=R+2}^{\infty} (i-\tfrac{1}{2})^r h^r \int_{(i-1)h}^{ih} f(x)\,dx = 0, \tag{27}$$

if $h = (b-a)/(R+1)$. If, therefore, we denote the above sums by $\epsilon(r, b)$ respectively we have, from (6),

$$\sum_{i=1}^{R+1} f_i \xi_i^r + \epsilon(r, b) = \mu_r + C(r, h) + S(r, h). \tag{28}$$

Applying now the previous method we introduce an additional error in the calculation of f_i, but this error is smaller than $\quad +\max|\epsilon(r, b)|\,\Sigma\,|u_{ir}|.$

The precise determination of the $\epsilon(r, b)$ for any given b would, of course, require a knowledge of the nature of the convergence in (26), i.e. some external knowledge about the distribution $f(x)$ which we are seeking to determine numerically. Unfortunately, such knowledge will in general not be available.

However, if b is chosen sufficiently large, the f_i determined for *different* values of b should all yield, by the method of §§ 2 and 3, approximations to the *same* probability integral $P(x)$ to within the errors of the respective remainder terms $S(r, h)$ and to within the errors introduced by (27). In practice, therefore, one would make an intelligent guess at the likely range of b and then test for internal consistency by comparing the probability integral tables obtained by varying b over this range. This method, which is illustrated in § 7 gives an idea of the accuracy to which the integral has been determined, but no rigorous statement on accuracy can be made without appealing to some *a priori* knowledge about $f(x)$. It is hoped to deal with this aspect more fully in the next paper.

7. The calculation of the χ-distribution for 10 degrees of freedom

As an illustration of the preceding section, we will now calculate the χ-distribution for 10 degrees of freedom. This distribution has high contact at either terminal and, although it is known to start at $x = \chi = 0$, we shall treat it as a distribution of double infinite range, i.e. we shall not make direct use of the information that $f(x) = 0$ for $x \leqslant 0$, and choose a *truncated* range $a \leqslant x \leqslant b$.

We have a mean of $\mu_1 = 3{\cdot}0843\,2776$, and the moments about the mean are given by†

$\mu_2' = 0{\cdot}486\,9223$, $\mu_3' = 0{\cdot}080\,6720$, $\mu_4' = 0{\cdot}713\,2999$, $\mu_5' = 0{\cdot}386\,6784$, $\mu_6' = 1{\cdot}810\,4865$.

* A formula for $S(r, h)$ when the range is infinite will be given in the second paper.

† These follow from the formulae for the moments about the origin which are ratios of Γ-functions (see, for example, Kendall, 1945, p. 55). Note that we have used μ and μ' for moments about the origin and the mean, respectively.

The standard deviation is $\sqrt{\mu_2'} = 0{\cdot}7$, and with seven group intervals available to cover the essential range we should choose h of the order of the standard deviation.* Our first attempt is, therefore, (a) $h = 0{\cdot}8$.

(a) If we make the mean of x the centre point of the innermost interval we have for the truncated range $a = \mu_1 - 3{\cdot}5 \times 0{\cdot}8 = \mu_1 - 2{\cdot}8$ and $b = \mu_1 + 2{\cdot}8$. For the standard variate X, the origin $X = 0$ will coincide with the mean of x and its range will be $-3{\cdot}5 \leqslant X \leqslant +3{\cdot}5$. Calculation of the moments (M_r) of X and substitution in the matrix (22) yields the following answers for the frequencies f_i:

$$f_1 = 0{\cdot}0005, \quad f_2 = 0{\cdot}03325, \quad f_3 = 0{\cdot}26266, \quad f_4 = 0{\cdot}42471,$$
$$f_5 = 0{\cdot}23196, \quad f_6 = 0{\cdot}04206, \quad f_7 = 0{\cdot}00487.$$

The calculated frequency (f_7) for the interval $\mu_1 + 2{\cdot}0 \leqslant x \leqslant \mu_1 + 2{\cdot}8$ is about 0·005, and its contribution to μ_6' about $0{\cdot}005\,(2{\cdot}4)^6 \sim 1$. Since this is an appreciable proportion of μ_6' it is unlikely that the frequencies beyond $b = \mu_1 + 2{\cdot}8$ when substituted in (27) can be neglected, i.e. b and h are too small.†

(b) Choosing therefore a larger h, we try $h = 1$. If we still keep the mean in the centre of the truncated range we have $a = \mu_1 - 3{\cdot}5 = -0{\cdot}42$ and $b = \mu_1 + 3{\cdot}5 = 6{\cdot}58$ (we know, of course, that $f(x) = 0$ for $x = 0$ so that our f_1 will really be the frequency for the interval $0 \leqslant x \leqslant 0{\cdot}085$). This time the standard variate is $X = x - \mu_1$ so that $M_r = \mu_r'$, and the above values can be substituted directly in the matrix (22) yielding the comparison of calculated χ-integral and 'exact' χ-integral as shown in Table 2.

Table 2. *Comparison of calculated and exact values of the χ-integral*

$X = \chi - \mu_1$	$P(x)$ exact	$P(x)$ calculated	Difference 10^{-5}
$-2{\cdot}5$	0·000 06	0·000 11	-5
$-1{\cdot}5$	0·009 29	0·008 93	36
$-0{\cdot}5$	0·244 66	0·244 75	-9
$+0{\cdot}5$	0·767 67	0·767 85	-18
$+1{\cdot}5$	0·979 02	0·978 88	14
$+2{\cdot}5$	0·999 45	0·999 47	-2
$+3{\cdot}5$	1·000 00	1·000 00	0

The maximum error is about 0·0004 and, again, the terminal f_i have a higher decimal accuracy. In practice, of course, the exact distribution would not be available for comparison. This time the terminal value f_7 is about 0·0005 and represents the frequency for the interval $\mu_1 + 2{\cdot}5 \leqslant x \leqslant \mu_1 + 3{\cdot}5$. Its contribution to μ_6' is about 0·4, thereby confirming that the previous grid of group intervals was too fine. To obtain further confirmation on the tail of the distribution, we determine a third set of f_i by shifting the grid of group intervals by 0·5 to the right, retaining the interval $h = 1$. This will make $a = \mu_1 - 3$ and $b = \mu_1 + 4$, i.e. $0{\cdot}08 \leqslant x \leqslant 7{\cdot}08$. For our standard variate X the origin will now coincide with $\mu_1 + 0{\cdot}5$.

* An unsuitable choice of h would, later, fail to satisfy the checks of internal consistency.

† Comparison with the exact χ-distribution shows that the maximum error in the above f_i is nevertheless not more than 0·005.

The values of the M_r are as follows:

$$M_1 = -0{\cdot}5, \quad M_2 = 0{\cdot}736\,9223, \quad M_3 = -0{\cdot}774\,7114,$$
$$M_4 = +1{\cdot}344\,8394, \quad M_5 = -1{\cdot}834\,794, \quad M_6 = 3{\cdot}595\,7606.$$

Substituting these in the matrix (22) we obtain the following values of f_i:

$$f_1 = 0{\cdot}000\,15, \quad f_2 = 0{\cdot}070\,12, \quad f_3 = 0{\cdot}444\,30, \quad f_4 = 0{\cdot}404\,21,$$
$$f_5 = 0{\cdot}076\,99, \quad f_6 = 0{\cdot}004\,19, \quad f_7 = 0{\cdot}000\,04.$$

The comparison of the progressive sums of the above f_i with the exact χ-integral is of similar accuracy to that in Table 2. The terminal frequency for $\mu_1 + 3 \leqslant x \leqslant \mu_1 + 4$ is 0·0005 with a contribution of about 0·03 to μ_6', indicating that we have now reached a satisfactory choice of b.

As a final check on the internal consistency we compare the answers obtained with the two last choices of group intervals by merging the tables of $P(x)$ to obtain one table at interval 0·5. This is set out in Table 3. The differences provide a fair check on the internal consistency to about 3-decimal accuracy of the two separate tables. If a more reliable check is desired, three or even four separate tables may be computed, all at the same group interval h and merged in the above manner to form a single table at interval $\frac{1}{3}h$ or $\frac{1}{4}h$. This procedure has the added advantage that interpolation difficulties at the wide interval of h are being avoided.

Table 3. $x = \chi$ *for* 10 *degrees of freedom. Calculated table of* $P(x)$ *obtained from two separate grids of group intervals* ($h = 1$)

$x-\mu_1$	$P(x)$			
−2·5	0·0001			
		1		
−2·0	0·0002		86	
		87		442
−1·5	0·0089		528	
		615		601
−1·0	0·0704		1129	
		1744		−174
−0·5	0·2448		955	
		2699		−1122
0	0·5147		−167	
		2532		−855
0·5	0·7679		−1022	
		1510		112
1·0	0·9189		−910	
		600		480
1·5	0·9789		−430	
		170		296
2·0	0·9959		−134	
		36		103
2·5	0·9995		−31	
		5		
3·0	1·0000			

8. The special case of symmetrical distributions; the normal integral

By placing the origin of the standard variate X at the mean of a symmetrical distribution we obviously have $f_1 = f_{R+1}$, $f_2 = f_R$, etc., i.e. the number of unknowns is halved. On the other hand, the odd moments contribute the meaningless equations

$$\Sigma f_i(\Xi_i^r + \Xi_{R+2-i}^r) = \Sigma f_i \times 0 = 0.$$

With the number of unknowns and equations halved and with even moments only retained, it is necessary to work out a new matrix ($\hat{V}_R$ say) based on even-order moments only. In practice the important values of R are $R = 4, 6, 8$ and 10, and we are giving below the inverse matrix $\hat{V}_8^{-1}$ (for $R = 8$) having rank 5 (as there are five equations corresponding to μ_0, μ_2, μ_4, μ_6 and μ_8):

i	f_i	$M_0 = 1$	M_2	M_4	M_6	M_8
1	f_1	0·000 3441	−0·001 7857	0·001 2153	−0·000 2315	0·000 0124
2	f_2	−0·003 9874	0·020 8333	−0·013 7153	0·002 3148	−0·000 0868
3	f_3	0·022 4151	−0·119 0476	0·071 1806	−0·008 1019	0·000 2480
4	f_4	−0·088 4281	0·500 0000	−0·143 4028	0·012 7315	−0·000 3472
5	f_5	0·569 6563	−0·400 0000	0·084 7222	−0·006 7130	0·000 1736

(29)

Working rule: Each f_i is obtained by forming the sums of five products using the five coefficients in the ith line and applying them to $M_0, \ldots, M_8$; e.g. $f_1 = 0{\cdot}000\,3441 M_0 - \ldots + 0{\cdot}000\,0124 M_8$.

As an example we compute the normal integral from its first five even moments, μ_0 to μ_8, choosing $h = 1$ and the standard variate X as normal deviate. Substituting, therefore, in the matrix (29) $M_0 = 1$, $M_2 = 1$, $M_4 = 3$, $M_6 = 15$ and $M_8 = 105$, we obtain the five f_i which in Table 4 have been progressively added to form the 'calculated normal integral' to be compared with the 'exact' one. The accuracy is remarkable, the maximum error being 15 in the 6th decimal.

Table 4. *Comparison of calculated normal integral with exact normal integral*

$x = X$	Exact $P(x)$	Calculated $P(x)$	Difference $\times 10^{-6}$
−4	0·000 032	0·000 034	− 2
−3	0·001 350	0·001 342	8
−2	0·022 750	0·022 765	−15
−1	0·158 655	0·158 643	12
0	0·500 000	0·500 000	0

With symmetrical distributions we cannot, of course, shift the grid of group intervals, as otherwise we would lose the symmetry relation between the f_i. If, therefore, intermediate values of $P(x)$ are required in order to ease subsequent interpolation, we can achieve this only by altering h. Merging the answers obtained from (say) three different h grids all centred at $x = 0$ (e.g. $h = 0{\cdot}9$, $1{\cdot}0$ and $1{\cdot}1$), we would *not* obtain a table of $P(x)$ at an *equidistant* interval. In the internal check we would, therefore, use divided differences.

9. Divergent or poorly convergent moments; the t-distribution for 10 degrees of freedom

Some variates with infinite range have distribution functions with low contact at $x = \infty$, i.e. the convergence in

$$\lim_{x \to \infty} f(x) = 0 \tag{30}$$

is slow, indeed, in some cases the moment μ_r is divergent for, say, $r \geqslant R'$.

As an example we have investigated the t-distribution for 10 degrees of freedom. Here we have $f(x) = c(1 + t^2/10)^{-5{\cdot}5}$ and hence $R' = 10$. In this case, therefore, R' is known *a priori*. If no such mathematical information is available, warning of low contact is given by the rapid

growth of the moments as $r \to R$, provided R is near to R'.* For our example for the t-distribution we find

$$\mu_2 = 1{\cdot}25, \quad \mu_4 = 6{\cdot}25, \quad \mu_6 = 78{\cdot}125, \quad \mu_8 = 2734{\cdot}375.$$

The difficulty with such distributions is that artificial truncation is not justified if the high-order, poorly convergent moments are to be used in equations (6). The remedy in such cases is the square variate transformation $y^2 = x$. Sometimes it may be necessary to use a higher power $y^k = x$. Obviously, if we were to take an equidistant interval for y, the group integral for x will *grow* with the square law, thereby absorbing the slowly convergent tail end of $f(x)$.

Now, obviously, the moments of y are simply related to those of x; we have

$$\int_0^\infty x^r f(x)\,dx = 2\int_0^\infty y^{2r} y f(y^2)\,dy, \tag{31}$$

or introducing the new distribution function $g(y) = 2yf(y^2)$, we have

$$\int_0^\infty x^r f(x)\,dx = \int_0^\infty y^{2r} g(y)\,dy. \tag{32}$$

Applying now the previous method to $g(y)$ it is further necessary to avoid using the poorly convergent high-order moments. In the case of the t-distribution, instead of taking $r = 0, 2, 4, 6$ and 8, we take the absolute moments† for $r = 0, 1, 2, 3$ and 4, which, according to (32), correspond to the even moment of $g(y)$. If only even moments about the origin are used in the determination of the f_i, the matrix (29) gives the appropriate inversion. Using $h = 0{\cdot}6$ for the y-group interval we substitute in (29):

$$M_0 = 1, \quad M_2 = 2{\cdot}401\,906, \quad M_4 = 9{\cdot}645\,062, \quad M_6 = 52{\cdot}952\,032 \quad \text{and} \quad M_8 = 372{\cdot}108\,863.$$

We thereby obtain five values of f_i $(i = 1, \ldots, 5)$ of the form

$$f_i = \int_{(5-i)h}^{(6-i)h} g(y)\,dy = \int_{(5-i)^2h^2}^{(6-i)^2h^2} f(x)\,dx. \tag{33}$$

The progressive sums of these are compared with the corresponding values of the exact t-integral in Table 5. Although the accuracy is lower than in the previous example it is satisfactory and very much better than we could have obtained without applying the transformation $y^2 = x$.

Table 5. *Comparison of calculated and exact values of the t-integral*

t	$P(t)$ exact	$P(t)$ calculated	Difference $\times 10^{-4}$
5·76	0·0001	0·0001	0
3·24	0·0044	0·0042	2
1·44	0·0902	0·0905	− 3
0·36	0·3613	0·3648	−35
0·00	0·5000	0·5000	0

* If R is much smaller than R', the present difficulty will not arise at all.

† We shall show in a second paper that, if the absolute moments of a distribution are not known, they can be obtained by interpolation between the values of $\log \mu_r$ for $r = 2, 4, 6, 8$, etc.; in fact, we shall give a general discussion of the interpolability of the logarithmic moment function for positive x.

10. LACK OF HIGH CONTACT AT THE START OF THE VARIATE $x=a$

We confine ourselves here to the most important case of lack of high contact at one terminal, say the start of the distribution, and assume, therefore, that there *is* high contact at *one end of the range*.

Without loss of generality we assume that $a = 0$, i.e. $x \geqslant 0$, and introduce the new variate $y^k = x$, $k \geqslant 2$. Whence we have

$$\int_0^b x^r f(x)\,dx = \int_0^{b^{1/k}} y^{kr} g(y)\,dy, \tag{34}$$

where $g(y) = ky^{k-1}f(y^k)$. Obviously $g(y)$ has, at least, contact of order $k-1$ at the start $y = 0$; further, if $f(x)$ has contact of order m at $x = b$, i.e. if $f(x) = O(x^{-m})$ at $x = b$, then $g(y) = O(y^{-(m-1)k-1})$. Hence there is high-order contact, of order $k-1$ and $(m-1)k+1$, respectively, at both ends of the range. The previous method is therefore applicable to $g(y)$ provided we can obtain its moments from those of $f(x)$. It is obvious from (34) that in order to obtain the ordinary moment of $g(y)$ we require to know the 'fractional' moments of $f(x)$, i.e. those corresponding to $r = j/k$ ($j = 0, 1, \ldots$). If the moments of $f(x)$ are only known for integer r the fractional moments will have to be obtained by interpolation of the logarithmic moment function $\log \mu_r$ which will be more fully discussed in the next paper.

REFERENCES

BARTLETT, M. S. (1937). *Proc. Roy. Soc.* A, **160**, 268.
FISHER, R. A. (1929). *Proc. Lond. Math. Soc.* (2), **30**, 199.
FISHER, R. A. (1930). *Proc. Roy. Soc.* A, **130**, 16.
GEARY, R. C. (1947). *Biometrika*, **34**, 68.
GEARY, R. C. & WORLLEDGE, J. P. G. (1947). *Biometrika*, **34**, 98.
HARTLEY, H. O. (1940). *Biometrika*, **31**, 249.
KENDALL, M. G. (1938). *J.R. statist. Soc.* **101**, 592.
KENDALL, M. G. (1945). *The Advanced Theory of Statistics*, **1**. London: C. Griffin and Co.
MILNE THOMSON, L. M. (1933). *The Calculus of Finite Differences.* London: Macmillan.
NAIR, U. S. (1936). *Biometrika*, **30**, 274.
NAYER, P. P. N. (1936). *Statist. Res. Mem.* **1**, 38.
NEYMAN, J. & PEARSON, E. S. (1931). *Bull. int. Acad. Cracovie*, A, p. 460.
SUKHATME, P. V. (1936). *Statist. Res. Mem.* **1**, 94.
Tables of the Incomplete Beta-Function, edited by Karl Pearson (1934). London: Biometrika Office.
WELCH, B. L. (1935). *Biometrika*, **27**, 145.
WELCH, B. L. (1936). *Statist. Res. Mem.* **1**, 1.

APPROXIMATION TO PERCENTAGE POINTS OF THE z-DISTRIBUTION

By A. H. CARTER, *King's College, Cambridge*

Tables have been published of the values of z for various percentage levels (20, 5, 1 and 0·1 %) for a range of given n_1, n_2 (Fisher & Yates, 1943, Table V). When n_1 or n_2 is outside the range of the tables, recourse must be had to approximate formulae (unless, of course, interpolation is sufficiently accurate) which will combine accuracy with facility of computation. One such formula, due to Fisher, with a modification suggested by Cochran (1940), is given at the foot of the above-mentioned tables. The purpose of this paper is to derive an alternative formula, no more difficult to compute, which will be shown to give consistently closer approximations to the true value of z for all except small n_1 or n_2.

Wishart (1947) has derived formulae for the exact cumulants of z, and also the well-known approximations to them when n_1, n_2 are large. The exact cumulants as far as κ_5 can be readily obtained arithmetically from tables of the Polygamma functions. Knowing the cumulants of the distribution, we may make use of the Cornish-Fisher normalization function method, based on Edgeworth's form of the Gram-Charlier type A series (Cornish & Fisher, 1937), to approximate to the percentage points. The method consists in writing z as an expansion in powers of a corresponding normal variate, ξ, the coefficients being functions of the cumulants of z, and assumes that κ_r is of order n^{1-r}, which is true for the z-distribution (Wishart, 1947, p. 172).

If z and ξ are expressed in standard measure (i.e. mean zero, standard deviation unity) we then derive

$$\frac{z-\mu_1'}{\sigma} = z' \sim \xi + \frac{\kappa_3}{\sigma^3}\frac{\xi^2-1}{6} + \frac{\kappa_4}{\sigma^4}\frac{\xi^3-3\xi}{24} - \frac{\kappa_3^2}{\sigma^6}\frac{2\xi^3-5\xi}{36},$$

correct to order n^{-1} for z'. This gives

$$\textit{Formula (a)}: \qquad z \sim \mu_1' + \sigma\xi + \frac{\kappa_3}{\sigma^2}\frac{\xi^2-1}{6} + \frac{\kappa_4}{\sigma^3}\frac{\xi^3-3\xi}{24} - \frac{\kappa_3^2}{\sigma^5}\frac{2\xi^3-5\xi}{36}, \tag{1}$$

correct to order $n^{-\frac{3}{2}}$ (since $\sigma = O(n^{-\frac{1}{2}})$), where $\mu_1'(=k_1)$, $\sigma^2(=\kappa_2)$, κ_3, κ_4 are cumulants of the z-distribution. The ξ-coefficients may be readily computed: e.g. for the 5 % level, substitute $\xi = 1\cdot64485$. Table 2 gives the values, for the 20, 5, 1 and 0·1 % levels, of the coefficients required in applying the formula. The quantities μ_1', σ, κ_3, κ_4 depend of course on n_1 and n_2, and may be evaluated in any particular case, whence substitution in (1) gives the appropriate value of z. Since $|z_{1-P}(n_1, n_2)| = |z_P(n_2, n_1)|$, where z_P is the value of z corresponding to probability P, to find the percentage points for the 'negative tail', i.e. 80, 95, 99 and 99·9 %, we may simply interchange n_1 and n_2. This has the effect of changing the sign of the *odd* cumulants, so that in (1) we write $-\mu_1'$ and $-\kappa_3$ for μ_1' and κ_3.

Formula (*a*), being an approximation to order $n^{-\frac{3}{2}}$, may be expected to give reliable results when n_1 and n_2 are both large. For the 1 % point, for example, we find $z(6, 12) = 0\cdot7843$ (true value 0·7864), whereas $z(24, 60) = 0\cdot3744$ (true value 0·3746). Some further results for (6, 12), (6, 60), and (24, 60) are shown in Table 1 (*a*).

In practice, some labour is involved in applying formula (*a*), even if polygamma tables are available. The Fisher-Cochran formula, derived by the normalization function method, is a simple working approximation, valid for large n_1, n_2, in which the exact cumulants are replaced by their approximations in terms of inverse powers of n_1 and n_2.

Table 1. *Comparison of approximations to the percentage points of z*

Formula (*b*): Existing formula (Fisher-Cochran).
Formula (*c*): New formula.

Percentage level	$n_1, n_2 \rightarrow$	6, 12	6, 60	24, 60	20, 36	20, 100	36, 60	24, 24	36, 36
20	Formula (*b*)	0·2687	0·1901	0·1335	0·1577	0·1287	0·1212	0·1741	0·1415
	Formula (*c*)	0·2733	0·2020	0·1340	0·1580	0·1298	0·1213	0·1740	0·1415
	True *z*	0·2706	0·1965	0·1338	0·1579	0·1294	0·1213	0·1740	0·1415
5	Formula (*b*)	0·5507	0·3990	0·2650	0·3128	0·2573	0·2390	0·3426	0·2778
	Formula (*c*)	0·5501	0·4100	0·2654	0·3129	0·2586	0·2391	0·3426	0·2778
	True *z*	0·5487	0·4064	0·2654	0·3129	0·2583	0·2391	0·3425	0·2778
1	Formula (*b*)	0·7992	0·5646	0·3746	0·4441	0·3619	0·3385	0·4894	0·3955
	Formula (*c*)	0·7886	0·5698	0·3746	0·4435	0·3629	0·3384	0·4893	0·3955
	True *z*	0·7864	0·5687	0·3746	0·4435	0·3630	0·3384	0·4890	0·3954
0·1	Formula (*b*)	1·1074	0·7474	0·4963	0·5928	0·4755	0·4503	0·6602	0·5307
	Formula (*c*)	1·0693	0·7372	0·4954	0·5906	0·4756	0·4498	0·6595	0·5304
	True *z*	1·0628	0·7377	0·4955	0·5905	0·4760	0·4498	0·6589	0·5302

Percentage level	$n_1, n_2 \rightarrow$	12, 6	60, 6	60, 24	36, 20	100, 20	60, 36
20	Formula (*b*)	0·3509	0·3346	0·1566	0·1783	0·1656	0·1314
	Formula (*c*)	0·3506	0·3408	0·1569	0·1785	0·1665	0·1315
	True *z*	0·3510	0·3388	0·1568	0·1784	0·1661	0·1315
5	Formula (*b*)	0·6884	0·6435	0·3047	0·3483	0·3208	0·2566
	Formula (*c*)	0·7001	0·6706	0·3060	0·3493	0·3236	0·2570
	True *z*	0·6931	0·6596	0·3055	0·3488	0·3227	0·2568
1	Formula (*b*)	1·0120	0·9444	0·4368	0·4995	0·4615	0·3661
	Formula (*c*)	1·0370	0·9956	0·4391	0·5016	0·4662	0·3667
	True *z*	1·0218	0·9770	0·4385	0·5009	0·4666	0·3666
0·1	Formula (*b*)	1·4352	1·3340	0·5930	0·6789	0·6303	0·4932
	Formula (*c*)	1·4681	1·4155	0·5965	0·6820	0·6375	0·4942
	True *z*	1·4449	1·3929	0·5962	0·6814	0·6371	0·4940

Table 1 (*a*). *Some values of z from formula* (*a*) (*exact cumulant formula*)

(For corresponding true values, see Table 1)

% \ n_1, n_2	6, 12	6, 60	24, 60	12, 6	60, 6	60, 24
20	0·2699	0·1998	0·1338	0·3499	0·3335	0·1567
5	0·5457	0·4022	0·2652	0·6958	0·6627	0·3057
1	0·7843	0·5640	0·3744	1·0295	0·9854	0·4388
0·1	1·0684	0·7433	0·4956	1·4592	1·4026	0·5966

The cumulant function of z is

$$K(z) = \tfrac{1}{2}it\log\frac{n_2}{n_1}+\log\Gamma\left(\frac{n_1+it}{2}\right)+\log\Gamma\left(\frac{n_2-it}{2}\right)-\log\Gamma\left(\frac{n_1}{2}\right)-\log\Gamma\left(\frac{n_2}{2}\right),$$

and the cumulants are obtainable by differentiating this successively with respect to (it), at each stage putting $t=0$.

Since
$$\left[\frac{d^r}{d(it)^r}\log\Gamma\left(\frac{n\pm it}{2}\right)\right]_{t=0} = (\pm 1)^r\frac{d^r}{dn^r}\log\Gamma\left(\frac{n}{2}\right),$$

and $\log\Gamma(\frac{1}{2}n)$ can be expanded by Stirling's theorem in inverse powers of n, the cumulants may also readily be expressed in inverse powers of n_1, n_2; and, when n_1, n_2 are reasonably large, the first few terms only in the expansions will give sufficiently close approximations to them. In fact, writing

$$\frac{1}{n_1}+\frac{1}{n_2}=s,\quad \frac{1}{n_1}-\frac{1}{n_2}=d,$$

it has been shown that

$$\left.\begin{aligned}\kappa_1 &= \mu_1' = -\tfrac{1}{2}d-6sd+O(n^{-4}),\\ \kappa_2 &= \sigma^2 = \tfrac{1}{2}s+\tfrac{1}{4}(s^2+d^2)+O(n^{-3}),\\ \kappa_3 &= -\tfrac{1}{2}sd+O(n^{-3}),\\ \kappa_4 &= \tfrac{1}{4}s(s^2+3d^2)+O(n^{-4}),\\ \kappa_r &= O(n^{1-r})\quad (r>1).\end{aligned}\right\}\tag{A}$$

Formula (a) will now have an extra term, since we take as our 'working variance' of z not its exact value, but its approximation to order n^{-1} from (A), i.e. $\frac{1}{2}s$. In the notation of Kendall (1945)

$$l_2 = \kappa_2/\tfrac{1}{2}s-1\sim\tfrac{1}{2}(s+d^2/s).$$

We then obtain

$$z-\mu_1'\sim\sqrt{\left(\frac{s}{2}\right)}\,\xi-\frac{d}{6}(\xi^2-1)+\sqrt{\frac{s}{2}}\left\{\frac{s}{24}(\xi^3+3\xi)+\frac{d^2}{72s}(\xi^3+11\xi)\right\}\tag{2}$$

$$=\frac{\xi}{\sqrt{2/s}}\left\{1+\frac{1}{2/s}\frac{\xi^2+3}{12}\right\}-\frac{d}{6}(\xi^2-1)+\frac{d^2}{144}\sqrt{\frac{2}{s}}(\xi^3+11\xi)$$

or
$$z-\mu_1'\sim\frac{\xi}{\sqrt{(h-\lambda)}}-\frac{d}{6}(\xi^2-1),\tag{3}$$

where
$$h=\frac{2}{s},\quad \lambda=\frac{\xi^2+3}{6},$$

provided $\frac{1}{144}d^2\sqrt{(2/s)}(\xi^3+11\xi)$ may be neglected (which will be the case for small d). Inserting the approximation to μ_1' from (A), i.e. $-\frac{1}{2}d$,

$$z\sim\frac{\xi}{\sqrt{(h-\lambda)}}-\frac{d}{6}(\xi^2+2),\tag{3a}$$

the Fisher-Cochran formula, which has, in fact, been found to give a fairly close approximation to the true z for n_1, n_2 both reasonably large. It may be noted that if n_1, n_2 are not very large, an improvement will be effected by including the second term in the estimate of the mean (κ_1), i.e. from (A) by adding $-\frac{1}{6}sd$.

For $(n_1, n_2) = (6, 12)$ this correction is $-0{\cdot}00347$, and for $(24, 60)$, the correction is $-0{\cdot}00024$. Inserting this improved approximation to μ_1' in (3) we have

Formula (b):
$$z\sim\frac{\xi}{\sqrt{(h-\lambda)}}-\frac{d}{6}(\xi^2+2+s).\tag{3b}$$

As pointed out by Wishart (1947, p. 179), an approximation to the value of any $\kappa_r (r > 1)$ obtained by considering its leading term only, will be improved by writing $1/(n_1-1)$ and $1/(n_2-1)$ in place of $1/n_1$ and $1/n_2$. For, by Stirling's expansion of a factorial,

$$\begin{aligned}
\log \Gamma\left(\frac{n}{2}\right) &= \frac{n-1}{2}\log\frac{n}{2}-\frac{n}{2}+\tfrac{1}{2}\log 2\pi+\frac{1}{6n}+O(n^{-3}),\\
\frac{d}{dn}\log \Gamma\left(\frac{n}{2}\right) &= \tfrac{1}{2}\log n-\frac{1}{2n}-\frac{1}{6n^2}+O(n^{-4}),\\
\frac{d^2}{dn^2}\log \Gamma\left(\frac{n}{2}\right) &= \frac{1}{2n}+\frac{1}{2n^2}+\frac{1}{3n^3}+O(n^{-5})\\
&= \frac{1}{2(n-1)}+O(n^{-3}),\\
\frac{d^3}{dn^3}\log \Gamma\left(\frac{n}{2}\right) &= -\frac{1}{2(n-1)^2}+O(n^{-4}),
\end{aligned}$$

and so on.

Table 2. *ξ-coefficients required in applying formula (a)*

	20 %	5 %	1 %	0·1 %
ξ	0·84162	1·64485	2·32635	3·09023
$\frac{1}{6}(\xi^2-1)$	−0·04861	0·28426	0·73532	1·42492
$\frac{1}{24}(\xi^3-3\xi)$	−0·08036	−0·02018	0·23379	0·84332
$\frac{1}{36}(2\xi^3-5\xi)$	−0·08377	0·01878	0·37634	1·21026

Thus writing $$s' = \frac{1}{n_1-1}+\frac{1}{n_2-1}, \quad d' = \frac{1}{n_1-1}-\frac{1}{n_2-1},$$

we might expect a better approximation to z to be obtained, corresponding to that of Fisher and Cochran, if we use s' and d' instead of s and d.

Corresponding to equations (A) we have

$$\left.\begin{aligned}
\kappa_2 &= \tfrac{1}{2}s'-\tfrac{1}{24}s'(s'^2+3d'^2)+O(n^{-4}),\\
\kappa_3 &= -\tfrac{1}{2}s'd'+O(n^{-4}),\\
\kappa_4 &= \tfrac{1}{4}s'(s'^2+3d'^2)+O(n^{-5}),\\
\kappa_r &= O(n^{1-r}) \quad (r>1).
\end{aligned}\right\} \tag{B}$$

For the mean, however, $$\mu_1' = \kappa_1 = -\tfrac{1}{2}d-\tfrac{1}{6}sd+O(n^{-4}), \tag{4}$$
$$= -\tfrac{1}{2}d'+\tfrac{1}{3}s'd'+O(n^{-3}), \tag{4a}$$

If n^{-3} is *not* negligible (relative to the degree of accuracy desired) μ_1' should therefore be left in the form $-\frac{1}{2}d-\frac{1}{6}sd$.

Proceeding as before, we obtain

$$z-\mu_1' \sim \sqrt{\left(\frac{s'}{2}\right)}\,\xi-\frac{d'}{6}(\xi^2-1)+s'\sqrt{\left(\frac{s'}{2}\right)}\frac{\xi^3-3\xi}{24}+d'^2\sqrt{\left(\frac{2}{s'}\right)}\frac{\xi^3-7\xi}{144}, \tag{5}$$

whence $$z-\mu_1' \sim \frac{\xi}{\sqrt{(h'-\lambda')}}-\frac{d'}{6}(\xi^2-1), \tag{6}$$

where $$h' = 2/s', \quad \lambda' = \tfrac{1}{6}(\xi^2-3).$$

Since this is based in the first place on more accurate approximations to the cumulants κ_3 and κ_4, and since the term omitted from (5) in deriving (6) (i.e. $\frac{1}{144}d'^2\sqrt{(2/s')}(\xi^3-7\xi)$, is evidently numerically less than the corresponding term omitted in obtaining the Fisher-Cochran formula (i.e. $\frac{1}{144}d^2\sqrt{(2/s)}(\xi^3+11\xi)$, formula (6) might be expected to give an improved approximation to z. In fact, however, it does not, and the reason is not far to seek.

Consider the expansion of $\xi/\sqrt{(h-\lambda)}$ in both cases:

$$\frac{\xi}{\sqrt{(h-\lambda)}}=\frac{\xi}{\sqrt{h}}\left(1+\frac{\lambda}{2h}+\frac{3\lambda^2}{4h^2}+\dots\right),$$

where the terms are decreasing in magnitude (since $1/h=\frac{1}{2}s=O(n^{-1})$). Hence the error in neglecting all terms after the second will be approximately of the order of the third term. Now in the Fisher-Cochran approximation this term, $+\frac{3\lambda^2\xi}{4h^2\sqrt{h}}$, has the same sign as the omitted term $\frac{1}{144}d^2\sqrt{(2/s)}(\xi^3+11\xi)$ (both being of the same sign as ξ), so that the extra terms included will tend to compensate for the term omitted. In obtaining (6) from (5), on the other hand, the term omitted, $\frac{1}{144}d'^2\sqrt{(2/s')}(\xi^3-7\xi)$, will be of *opposite* sign to ξ when $|\xi|<\sqrt{7}$, corresponding to a probability of about 0·004: so that for most percentage levels encountered in practice, the error in (5) is increased in (6).

A better formula is obtained from (5) as

$$z-\mu_1'\sim\frac{\xi\sqrt{(h'+\lambda')}}{h'}-\frac{d'}{6}(\xi^2-1), \tag{7}$$

where h' and λ' are as in (6).

Expanding $\xi\sqrt{(h'+\lambda')}/h'$, it is found that the third term is now of *opposite* sign to ξ, and hence the extra terms contained in the expansion will tend to compensate for the term omitted. Since s' and d' require to be calculated in applying this formula, it is desirable to write μ_1' in the form (4*a*) (provided we can neglect quantities of order n^{-3}). This gives

Formula (*c*): $$z\sim\frac{\xi\sqrt{(h'+\lambda')}}{h'}-\frac{d'}{6}(\xi^2+2-2s'). \tag{7a}$$

Collecting the results, we have the three approximate formulae:

Formula (*a*) (exact cumulant method):

$$z\sim\kappa_1+\sigma\xi+\frac{\kappa_3}{\sigma^2}\frac{\xi^2-1}{6}+\frac{\kappa_4}{\sigma^3}\frac{\xi^3-3\xi}{24}-\frac{\kappa_3^2}{\sigma^5}\frac{2\xi^3-5\xi}{36}.$$

Formula (*b*) (Fisher-Cochran formula):

$$z\sim\frac{\xi}{\sqrt{(h-\lambda)}}-\frac{d}{6}(\xi^2+2+s),$$

where $$s=\frac{1}{n_1}+\frac{1}{n_2},\quad d=\frac{1}{n_1}-\frac{1}{n_2},\quad h=\frac{2}{s},\quad \lambda=\frac{\xi^2+3}{6}.$$

Formula (*c*) (new formula): $$z\sim\frac{\xi\sqrt{(h'+\lambda')}}{h'}-\frac{d'}{6}(\xi^2+2-2s'),$$

where $$s'=\frac{1}{n_1-1}+\frac{1}{n_2-1},\quad d'=\frac{1}{n_1-1}-\frac{1}{n_2-1},\quad h'=\frac{2}{s'},\quad \lambda'=\frac{\xi^2-3}{6}.$$

When n_1, n_2 are large (and not too different), $\frac{1}{6}sd$ is negligible and formula (b) becomes the formula more generally quoted

$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \left(\frac{1}{n_1} - \frac{1}{n_2}\right)\frac{\xi^2+2}{6}$$

which may be written

$$z \sim \frac{\xi}{\sqrt{(h-\lambda)}} - \left(\frac{1}{n_1} - \frac{1}{n_2}\right)(\lambda - \tfrac{1}{6}).$$

Similarly, for sufficiently large n_1, n_2, $\frac{1}{3}s'd'$ may be neglected and formula (c) becomes

$$z \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)\frac{\xi^2+2}{6},$$

or

$$z \sim \frac{\xi\sqrt{(h'+\lambda')}}{h'} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)(\lambda' + \tfrac{5}{6}).$$

It is to be noted, however, that since $\frac{1}{3}s'd'$ is approximately twice $\frac{1}{6}sd$, more care must be exercised in deciding to neglect it. For example, when (n_1, n_2) is (20, 100), $\frac{1}{3}s'd' = 0{\cdot}0009$, and for (24, 60), its value is 0·0005.

For purposes of comparison, values of z have been computed from formulae (b) and (c), for the four common percentage levels, over a fairly wide range of n_1, n_2. They are shown in Table 1, together with the corresponding true values of z. The latter were obtained where possible from the tables of Fisher and Yates: elsewhere by inverse interpolation in *Tables of the Incomplete Beta-Function* followed by a logarithmic transformation. Such values are in error by not more than 0·0001. It will be seen that neither formula yields very accurate results when n_1 or n_2 is as small as 6, though even here the new formula is rather better with the single exception of $n_1 = 12$, $n_2 = 6$. In actual practice, however, we are concerned with large values of n_1, n_2, beyond the range of the published tables. Considering only those cases where n_1 and n_2 are both greater than 20, it is seen that formula (c) gives a consistently closer approximation than does formula (b) for both the positive and the negative tails, and for all the percentage levels investigated, though its relative gain in accuracy is greatest at the 1 and 0·1 % levels. It may be noted, in fact, that in no case considered having n_1 and n_2 greater than 20, is the error more than 9 in the fourth decimal place, i.e. it appears that for all except small n_1, n_2, this formula will give an approximation to z correct to within 0·001.

In conclusion, therefore, it is recommended that formula (c) be adopted for general use, since it is no more difficult to compute, and is more accurate, than the existing formula. Dropping the dashes we have the formula

$$z \sim \frac{\xi\sqrt{(h+\lambda)}}{h} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)\left(\lambda + \frac{5}{6} - \frac{s}{3}\right),$$

where

$$s = \frac{1}{n_1-1} + \frac{1}{n_2-1}, \quad h = \frac{2}{s}, \quad \lambda = \frac{\xi^2-3}{6}$$

or, if $\frac{1}{3}\left\{\frac{1}{(n_1-1)^2} - \frac{1}{(n_2-1)^2}\right\}$ may be neglected,

$$z \sim \frac{\xi\sqrt{(h+\lambda)}}{h} - \left(\frac{1}{n_1-1} - \frac{1}{n_2-1}\right)(\lambda + \tfrac{5}{6}).$$

The values of ξ and λ for the four percentage levels are:

	20 %	5 %	1 %	0·1 %
ξ	0·8416	1·6449	2·3263	3·0902
λ	−0·3819	0·0491	0·4020	1·0916

My thanks are due to Dr J. Wishart, whose suggestion was the basis of this paper.

REFERENCES

Cochran, W. G. (1940). Note on an approximative formula for significance levels of z. *Ann. Math. Statist.* **11**, 93.

Cornish, E. A. & Fisher, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* **4**, 307.

Fisher, R. A. & Yates, F. (1943). *Statistical Tables*, 2nd ed. Oliver and Boyd.

Kendall, M. G. (1945). *Advanced Theory of Statistics*, **1**, 156, 2nd ed. C. Griffin and Co.

Wishart, J. (1947). The cumulants of the z and of the logarithmic χ^2 and t distributions. *Biometrika*, **34**, 170.

MISCELLANEA

Note on the cumulants of Fisher's z-distribution

By LEO A. AROIAN, *Hunter College*

In a recent article Dr J. Wishart (1947) stated: 'Explicit expressions for the exact cumulants of Fisher's z-distribution do not appear ever to have been published.' Fisher's z-distribution and the related Snedecor's F-distribution formed a part of my doctor's thesis and rather full results concerning the cumulants of the z-distribution and other properties of the distribution were published in the *Annals of Mathematical Statistics* (Aroian, 1941) some time ago.* I should like to take this opportunity of adding certain comments on the Gram-Charlier Type A approximation to the z-distribution and the type III approximation to the F-distribution.

To obtain the cumulants of the z-distribution I expanded the moment generating function $M_z(\theta)$ in powers of θ and found $\lambda_{k:z}$, the kth semi-variant (or cumulant) of z as the coefficient of $\theta^k/k!$. The exact results correspond with Wishart's formulae (9) to (15), although given in a different form, and need not be repeated here. In addition, asymptotic formulae for $\lambda_{k:z}$, n_1 and n_2 large, were derived by means of the Euler-Maclaurin sum formula. Furthermore, another type of formula could have been given for n_1 small but n_2 large, merely by expanding that part of $\lambda_{k:z}$ in which n_2 occurs by the Euler-Maclaurin sum formula. The special cases for the logarithmic χ^2, the logarithmic t, and the logarithmic normal probability functions follow by substituting the proper limiting values of n_1 and n_2.

In my previous paper I was overcautious concerning the type A approximation to the z-distribution. Actually the method is fairly accurate although tedious. Taking

$$F(t) = \phi(t) + A_3\phi^{\text{iii}}(t) + A_4\phi^{\text{iv}}(t), \int_{t_0}^{\infty} F(t)\,dt = \eta,$$

we have

$$\int_{t_0}^{\infty} F(t)\,dt = \int_{t_0}^{\infty}\phi(t)\,dt + \phi(t)\{-A_3(t_0^2-1) + A_4(t_0^3-3t_0)\},$$

where η is usually 0·10, 0·05, 0·025, 0·01, etc. As an example take $n_1 = 24$, $n_2 = 60$; then

$$\lambda_{1:z} = -0{\cdot}0127429, \quad \sigma_z = 0{\cdot}173779, \quad \lambda_{3:z} = -0{\cdot}0007998, \quad \lambda_{4:z} = 0{\cdot}0000867,$$

$$A_3 = \frac{-\lambda_{3:z}}{3!\,\sigma_z^3} = 0{\cdot}025345, \quad A_4 = \frac{\lambda_{4:z}}{4!\,\sigma_z^4} = 0{\cdot}00396.$$

t_0 for $\eta = 0{\cdot}05$ is $1{\cdot}60094$, $z_{0{\cdot}05} = 0{\cdot}26547$ against the accurate value of 0·26534. For the 1 % point $t_0 = 2{\cdot}2338$, $z_{0{\cdot}01} = 0{\cdot}3754$ against the accurate value of 0·3746. When $n_1 = n_2 = 24$, $z_{0{\cdot}05} = 0{\cdot}3423$ against the accurate value of 0·3425.

The type III approximation to the F-distribution is of some interest since for n_1 moderate and n_2 large, n_1F tends to be distributed as χ^2 with n_1 degrees of freedom. Since

$$\text{Mean } F = \bar{F} = \frac{n_2}{n_2-2}; \quad \sigma_F = \frac{n_2}{n_2-2}\sqrt{\frac{2(n_1+n_2-2)}{n_1(n_2-4)}},$$

$$\alpha_{3:F} = \frac{4(2n_1+n_2-2)}{n_1(n_2-6)}\sqrt{\frac{n_1(n_2-4)}{2(n_1+n_2-2)}}, \quad \alpha_{3:F} = \sqrt{(\beta_{1:F})},$$

we find the 5, 1 or 0·1 % points for F by using

$$F_{0{\cdot}05} = \bar{F} + \sigma_F(1{\cdot}64485 + 0{\cdot}28392\alpha_{3:F} - 0{\cdot}04902\alpha_{3:F}^2),$$

$$F_{0{\cdot}01} = \bar{F} + \sigma_F(2{\cdot}32635 + 0{\cdot}73330\alpha_{3:F} - 0{\cdot}024957\alpha_{3:F}^2),$$

$$F_{0{\cdot}001} = \bar{F} + \sigma_F(3{\cdot}0903 + 1{\cdot}4190\alpha_{3:F} + 0{\cdot}05667\alpha_{3:F}^2).$$

* [Both Dr Wishart, as author, and myself as editor regret that owing to wartime preoccupation the publication of Dr Aroian's 1941 paper was overlooked. E.S.P.]

These formulae for the levels of significance of the χ^2 distribution are from a previous paper (Aroian, 1943). For $n_1 = 24$, $n_2 = 60$, $F_{0.05}$ by this approximation is 1·709 compared with the accurate value of 1·700. For $n_1 = 24$, $n_2 = 100$, $F_{0.05}$ by this approximation is 1·631 against the accurate value of 1·627. For $n_1 = n_2 = 100$, $F_{0.05}$ by this approximation is 1·394 as compared with the accurate value of 1·392. While these results are not too poor, they are not so accurate as the well-known formulae of Cochran-Fisher or of E. Paulson (1942) which, for large values of n_1 and n_2, generally give 4 significant figures.

REFERENCES

Aroian, L. A. (1941). *Ann. Math. Statist.* **12**, 429.
Aroian, L. A. (1943). *Ann. Math. Statist.* **14**, 93.
Paulson, E. (1942). *Ann. Math. Statist.* **13**, 233.
Wishart, J. (1947). *Biometrika*, **34**, 170.

A note on the mean deviation from the median

By K. R. NAIR

For samples drawn from a normal universe, Godwin (1945) obtained the sampling distribution of the mean deviation when the individual deviations are measured from the sample *mean*. It is well known that the mean deviation is least when it is measured from the sample *median*.* Let us refer to them as 'mean deviation from mean' and 'mean deviation from median' respectively, and use the letters m and m' to denote their sample estimates.

The exact sampling distribution of m being now known and its probability integral tabulated, the question may well be asked what the distribution of m' is. Since $m' \leqslant m$, their expectations have the same relationship

$$E(m') \leqslant E(m). \tag{1}$$

For samples of n from a normal population with standard deviation, σ, $E(m') = f'_n\sigma$ and $E(m) = f_n\sigma$, where $f'_n \leqslant f_n$. For getting unbiased estimates of σ we should divide m' by f'_n and m by f_n. What we are now interested to know is which of the two estimates has a smaller standard error. In the case of m, it has been shown by Helmert (1876) and Fisher (1920) that

$$f_n = \sqrt{\frac{2(n-1)}{n\pi}}, \tag{2}$$

and

$$\text{S.E. of } \left(\frac{m}{f_n}\right) = \frac{\sigma}{\sqrt{n}}\sqrt{\left[\left\{\frac{\pi}{2} + \sqrt{[n(n-2)]} - n + \sin^{-1}\frac{1}{n-1}\right\}\right]}. \tag{3}$$

In the case of m', we neither know f'_n nor the standard error of (m'/f'_n) for samples of size n.

It is obvious that when n is very large, the mean and median will differ very little from one another and hence $m' \to m$ and $f'_n \to f_n$. It is interesting to note that, at the other end of the scale, namely, when $n = 2$, m and m' are identical, and equal to one-half the sample range.

To discover any real difference that may exist between the standard errors of (m/f_n) and (m'/f'_n), which is the same as determining the difference between the coefficients of variation of m and m', we must consider samples of size greater than 2.

(i) Let us take $n = 3$, and let x_1, x_2, x_3 be the observed values arranged in order of ascending magnitude. We at once find that

$$m' = \tfrac{1}{3}(x_3 - x_1). \tag{4}$$

The distribution of m' for samples of 3 is therefore derivable from that of the range. The probability integral of the range has been tabulated by Pearson & Hartley (1942) for $n = 2$ to 20. For our purpose it is necessary only to know the values of the mean range ($\bar{w}$) and the standard error of the range (σ_w)

* When n, the sample size, is an odd number, the sample median is by definition the value of the $\frac{1}{2}(n+1)$th ranked observation. When n is even, the sample median is conventionally taken as the mean of the $\frac{1}{2}n$th and $\frac{1}{2}(n+2)$th ranked values. The mean deviation from the median will have the same magnitude whatever value, between the $\frac{1}{2}n$th and $\frac{1}{2}(n+2)$th ranked values, the median takes, when n is even. No complication is therefore introduced by accepting the conventional definition of the

for samples of 3. This can be calculated, correct to six decimal places, from certain numerical values given by Pearson (1926). Using his figures,

$$\bar{w} = 1\cdot692568 \times \sigma, \tag{5}$$

$$\sigma_w = 0\cdot888368 \times \sigma. \tag{6}$$

The value of f'_n for sample of 3 is, therefore,

$$f'_3 = \tfrac{1}{3} \times 1\cdot692568 = 0\cdot56419,$$

and the standard error of m'/f'_3 is
$$\frac{0\cdot888368}{1\cdot692568}\sigma = 0\cdot52486\sigma, \tag{7}$$
correct to five decimal places.

The corresponding values for f_3 and standard error of (m/f_3) are obtained by putting $n = 3$ in equations (2) and (3) and are

$$f_3 = \sqrt{\frac{4}{3\pi}} = 0\cdot65147, \tag{8}$$

and
$$\text{S.E. of } (m/f_3) = \frac{\sigma}{\sqrt{3}}\sqrt{\left(\frac{2\pi}{3} + \sqrt{3} - 3\right)} = 0\cdot52486\sigma, \tag{9}$$
correct to five decimal places.

Although (9) can be evaluated to any number of decimal places, we are not in a position to bring (7) to a higher order of accuracy than five decimal places. It is very unlikely that (7) and (9) are absolutely identical, but we may safely conclude that they are practically the same.

(ii) We next come to samples of 4. If x_1, x_2, x_3, x_4 be the observations arranged in order of ascending magnitude, the mean deviation from median is given by

$$m' = \tfrac{1}{4}(x_4 + x_3 - x_2 - x_1). \tag{10}$$

The distribution of m' follows immediately from 'some order statistic distributions for samples of size 4' obtained by Walsh (1946) and is as follows:

$$p(m')\,dm' = \frac{12}{[\sqrt{(2\pi)}]^3} e^{-2m'^2} \left(\int_0^{4m'} e^{-\frac{1}{2}y^2} dy\right)^2 dm'. \tag{11}$$

The probability integral of m' is given by

$$P(m') = \int_0^{m'} p(m')\,dm' = \left(\int_0^{4m'} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}y^2} dy\right)^3 = \left(\frac{2}{\sqrt{(2\pi)}}\int_0^{2m'} e^{-\frac{1}{2}x^2} dx\right)^3. \tag{12}$$

The values of $P(m')$ given by (12) can easily be evaluated using the normal probability integral table and are given in cols. (3) and (6) of the table below, alongside corresponding values (given in cols. (2) and (5)) for the probability integral of the mean deviation (m) from the mean, for samples of 4, copied from Godwin's (1945) tables.

Table giving the probability integral of the mean deviation from (a) mean and (b) median for samples of four observations from a normal universe ($\sigma = 1$)

m (or m')	$P(m)$	$P(m')$	m (or m')	$P(m)$	$P(m')$
0·0	0·00000	0·00000	1·3	0·96758	0·97229
0·1	0·00333	0·00398	1·4	0·98229	0·98475
0·2	0·02534	0·03003	1·5	0·99073	0·99192
0·3	0·07879	0·09204	1·6	0·99534	0·99588
0·4	0·16693	0·19139	1·7	0·99775	0·99798
0·5	0·28345	0·31818	1·8	0·99895	0·99905
0·6	0·41552	0·45629	1·9	0·99953	0·99957
0·7	0·54836	0·58951	2·0	0·99980	0·99981
0·8	0·66934	0·70592	2·1	0·99992	0·99992
0·9	0·77040	0·79954	2·2	0·99997	0·99997
1·0	0·84860	0·86962	2·3	0·99999	0·99999
1·1	0·90502	0·91888	2·4	1·00000	1·00000
1·2	0·94321	0·95162			

We note that although m and m' have an infinite range from 0 to ∞, their probability integrals rapidly approach unity, this value being reached to five decimal place accuracy when $m(m') = 2{\cdot}4\sigma$. We can approximately work out the moments of the two distributions from the table above. The values of the mean and the standard deviation (applying Sheppard's correction for grouping) of m and m' so obtained are given below:

$$\left.\begin{array}{lll} \text{Mean:} & \overline{m} = 0{\cdot}690986\sigma, & \overline{m}' = 0{\cdot}663187\sigma, \\ \text{Standard deviation:} & \sigma_m = 0{\cdot}297015\sigma, & \sigma_{m'} = 0{\cdot}292979\sigma, \\ \text{Coefficient of variation:} & \sigma_m/\overline{m} = 0{\cdot}429842, & \sigma_{m'}/\overline{m}' = 0{\cdot}441775. \end{array}\right\} \tag{13}$$

The values of $\overline{m}$ and $\sigma_m/\overline{m}$ obtained from the exact formulae (2) and (3) are

$$\left.\begin{array}{r} \overline{m} = \sigma\sqrt{\left(\dfrac{3}{2\pi}\right)} = 0{\cdot}690988\sigma, \\ \sigma_m/\overline{m} = \frac{1}{2}\sqrt{(\frac{1}{2}\pi + \sin^{-1}\frac{1}{3} + 2\sqrt{2} - 4)} = 0{\cdot}429842, \end{array}\right\} \tag{14}$$

showing close agreement with the values given in (13) for the mean and coefficient of variation of m. We may therefore consider the mean and coefficient of variation of m', approximately evaluated in (13), to be of sufficient accuracy to warrant the conclusion that, for samples of size 4, the mean deviation from the mean leads to a more 'efficient' estimate of the population standard deviation than the mean deviation from the median. As the distribution of the latter is not known for $n > 4$, we are not in a position to say whether this conclusion holds good, in general, for all values of n.

In conclusion, it seems worth making the following point:

(a) if expressions for the expectation and variance of m' were available and tables of its probability integral worked out,

(b) if the efficiency of the m' estimate compared to the m estimate for $n > 4$ was not appreciably worse than for the case $n = 4$,

there would be strong practical grounds for using m' rather than m in view of greater simplicity in calculation. In both cases we must first arrange the observations in order of magnitude. Then if $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$, m' may be calculated from the formula

$$m' = \frac{1}{n}\{(x_{n-t+1} + x_{n-t+2} + \ldots + x_n) - (x_1 + x_2 + \ldots + x_t)\}, \tag{15}$$

where $t = \frac{1}{2}n$ or $\frac{1}{2}(n-1)$ according as n is even or odd.

For m, however, we must also calculate the arithmetic mean $\bar{x}$ and look for x_k and x_{k+1} between which $\bar{x}$ lies. Then m can be obtained from one of the three formulae

$$\left.\begin{array}{r} \dfrac{nm}{2k} = \bar{x} - \dfrac{x_1 + x_2 + \ldots + x_k}{k}, \\ \dfrac{nm}{2(n-k)} = \dfrac{x_{k+1} + \ldots + x_n}{n-k} - \bar{x}, \\ \dfrac{n^2 m}{2k(n-k)} = \dfrac{x_{k+1} + \ldots + x_n}{n-k} - \dfrac{x_1 + \ldots + x_k}{k}. \end{array}\right\} \tag{16}$$

This certainly involves a rather longer process.

It is interesting to note that m' becomes a special case of the measure of dispersion based on difference between the sums of the first and the last r observations (in order of magnitude) suggested by Jones (1946), the range, becoming another special case of the same measure, when $r = 1$.

REFERENCES

Fisher, R. A. (1920). *Mon. Not. R. Astr. Soc.* **80**, 758.
Godwin, H. J. (1945). *Biometrika*, **33**, 254.
Helmert, W. (1876). *Astr. Nachr.* **88**, no. 2096.
Jones, A. E. (1946). *Biometrika*, **33**, 274.
Pearson, E. S. (1926). *Biometrika*, **18**, 173.
Pearson, E. S. & Hartley, H. O. (1942). *Biometrika*, **32**, 301.
Walsh, J. E. (1946). *Ann. Math. Statist.* **17**, 246.

On the method of paired comparisons

By P. A. P. MORAN, *Institute of Statistics, Oxford University*

M. G. Kendall & B. Babington Smith (1940) have discussed the 'method of paired comparisons' for investigating preferences. Suppose we are given n objects $A, \ldots, K$, and an observer is asked to choose between every pair. If A is preferred to B we write $A \to B$. If the observer is not completely consistent, either because of his own inefficiency or because the objects are not really capable of being ranked in respect of the quality under consideration, he may make preferences of the type $A \to B \to C \to A$, and we call this an inconsistent or circular triad. Write d for the number of circular triads in a given experiment. Then Kendall & Babington Smith show that

$$\begin{aligned}\zeta &= 1 - \frac{24d}{n^3 - n} \quad (n \text{ odd}) \\ &= 1 - \frac{24d}{n^3 - 4n} \quad (n \text{ even})\end{aligned}$$

may be regarded as a 'coefficient of consistence' and lies between 0 and 1, being capable of attaining both these limits.

Now suppose that each comparison is made at random so that there are equal chances that $A \to B$ and $B \to A$. The distribution of d is then of interest. They calculate this distribution exactly for $n = 2, \ldots, 7$ and conjecture that its moments are given by

$$\mu_1' = \frac{1}{4}\binom{n}{3},$$

$$\mu_2 = \frac{3}{16}\binom{n}{3},$$

$$\mu_3 = -\frac{3}{32}\binom{n}{3}(n-4),$$

$$\mu_4 = \frac{3}{256}\binom{n}{3}\left\{9\binom{n-3}{3} + 39\binom{n-3}{2} + 9\binom{n-3}{1} + 7\right\},$$

these being polynomials in n which agree with their numerical calculations for $n = 2, \ldots, 7$. They also conjecture that the distribution tends to normality when n increases. In the present note we prove these statements.

Let the objects be numbered from 1 to n. Write $P_{ijk} = 1$ if the triad (i, j, k) is circular, and $P_{ijk} = 0$ if it is not. Then $d = \Sigma P_{ijk}$, the sum being taken over all such triads. Now by enumerating the various cases we see that $E(P_{ijk}) = \frac{1}{4}$ and so $\mu_1'(d) = E(\Sigma P_{ijk}) = \frac{1}{4}\binom{n}{3}$. Now consider $\mu_2'(d) = E[(\Sigma P_{ijk})^2]$. Consider the types of terms which results when we expand this. In the first place we have $\binom{n}{3}$ terms typified by P_{123}^2, and these contribute $\frac{1}{4}\binom{n}{3}$ to $\mu_2'(d)$. Similarly, we have terms typified by $P_{123}P_{145}$, $P_{123}P_{124}$ and $P_{123}P_{456}$, and the number of these are respectively $\frac{3}{2}\binom{n}{3}(n-3)(n-4)$, $3\binom{n}{3}(n-3)$ and $\binom{n}{3}\binom{n-3}{3}$, whilst their expectations are each $\frac{1}{16}$. It follows that

$$\mu_2'(d) = \frac{1}{16}\binom{n}{3}\left\{\binom{n}{3} + 3\right\},$$

and so

$$\mu_2(d) = \frac{3}{16}\binom{n}{3}.$$

The calculation of $\mu_3'(d)$ is a good deal more complicated. $\mu_3'(d) = E((\Sigma P_{ijk})^3)$, and on expanding we get 16 types of terms, typified by

$$\begin{array}{llll} P_{123}^2, & P_{123}^2 P_{145}, & P_{123}^2 P_{124}, & P_{123}^2 P_{456}, \\ P_{123} P_{145} P_{167}, & P_{123} P_{124} P_{125}, & P_{123} P_{145} P_{456}, & P_{123} P_{145} P_{146}, \\ P_{123} P_{124} P_{145}, & P_{123} P_{234} P_{134}, & P_{123} P_{345} P_{567}, & P_{123} P_{245} P_{346}, \\ P_{123} P_{145} P_{678}, & P_{123} P_{124} P_{567}, & P_{123} P_{456} P_{789}, & P_{123} P_{345} P_{245}. \end{array}$$

After some calculation we find the sum of the contributions of these to be

$$\frac{1}{2304}\binom{n}{3}\{n^6 - 6n^5 + 13n^4 + 42n^3 - 158n^2 - 108n + 864\}.$$

Reducing to the mean we get $\mu_3(d) = -\dfrac{3}{32}\dbinom{n}{3}(n-4)$.

The calculation of $\mu_4'(d)$ is a great deal more complicated, there being 85 terms which are not zero; we finally obtain, after lengthy calculations,

$$\mu_4' = \frac{1}{55296}\binom{n}{3}\{n^9 - 9n^8 + 33n^7 + 45n^6 - 582n^5 + 504n^4 + 5732n^3 - 10692n^2 - 30024n + 80352\},$$

and so
$$\mu_4 = \frac{1}{55296}\binom{n}{3}\{972n^3 + 972n^2 - 36936n + 80352\},$$

which reduces to the conjectured result.

We now prove that the distribution tends to normality. To do this, it is sufficient (Kendall, 1943, p. 110) to prove that

$$\frac{\mu_{2m}}{\mu_2^m} \to \frac{(2m)!}{2^m m!}, \quad \frac{\mu_{2m+1}}{\mu_2^{\frac{1}{2}(2m+1)}} \to 0, \quad \text{for} \quad m = 1, 2, \ldots.$$

Consider the second of these first. Write $Q_{ijk} = P_{ijk} - \frac{1}{4}$. Then

$$\mu_{2m+1}(d) = E[(\Sigma Q_{ijk})^{2m+1}].$$

It is clear that for any given m we could calculate $\mu_{2m+1}(d)$ given sufficient labour, by expanding this and considering the expectation of each type of term and calculating the number of times it occurs, which will be a polynomial in n. Now consider the various types of terms in the expansion of $(\Sigma Q_{ijk})^{2m+1}$. We classify these terms according to whether the Q's have common suffixes. Let $Q_{ijk} Q_{lmn} \ldots Q_{pqr}$ be a typical product in the expansion. If this can be separated into p groups of products of Q's such that different groups have no common suffixes whilst within each group the triads are connected to each other by having common points, we shall say such a product 'contains p groups'. Moreover, the number of times such a term occurs will be a polynomial in n whose order is equal to the number of distinct suffixes occurring in the product. If in a group a suffix only appears once, the inconsistency of the triad containing it is unaffected by the remainder of the group and the expectation of the product of Q's in that group will be zero. It follows that in all those terms which contribute something non-negative to μ_{2m+1}, none of the groups can contain a suffix which appears only once. Therefore, since all terms which contain more than m groups will have at least one group consisting of a single Q, the expectation of such terms will be zero. It follows that μ_{2m+1} is a polynomial in n, of degree $3m+1 = \left[\text{integral part of } \frac{3}{2}(2m+1)\right]$ at most, whose coefficients depend on m only. But $\mu_2(d)$ is of degree 3 in n and so

$$\frac{\mu_{2m+1}}{\mu_2^{\frac{1}{2}(m+1)}} = O\left(\frac{n^{3m+1}}{n^{3m+\frac{3}{2}}}\right) = O(n^{-\frac{1}{2}}).$$

Now consider μ_{2m}. This is a polynomial in n, and our aim is to find the order and coefficient of the term of largest order. In the first place we need only consider terms with m or less groups, for if a term has more than m groups, one at least will consist of a single Q and the expectation of the term will be zero. Moreover as before, in each term, the suffixes in each group must each occur at least twice in that group. The number of times each type of term occurs will be a polynomial in n of order equal to the total number of distinct suffixes in that term. As we shall show the leading term in $\mu_{2m}(d)$ to be of order $3m$, we can neglect terms whose frequency is less than this and therefore we can neglect all terms in which a suffix appears more than twice. Now consider a term with fewer than m groups and therefore containing a group of

order greater than two in the Q's. As no suffix can occur more than twice, no Q can occur more than once. Consider any Q_{ijk}, say, of this group. Then either the suffixes i, j, k are common to three other triads or one, i, say, is common to another triad and j, k common to a third. In either case evaluation of the expectation shows it to be zero. We can therefore restrict our attention to the case where there are m groups each containing two triads. Such groups can only be of the form Q^2_{123}, $Q_{123}Q_{124}$, $Q_{123}Q_{145}$ and the expectations of the two latter are zero whilst the expectation of Q^2_{123} is

$$\tfrac{1}{4}(1-\tfrac{1}{4})^2+\tfrac{3}{4}(\tfrac{1}{4})^2 = \tfrac{3}{32}.$$

The number of groups is m and the number of ways of choosing m such distinct pairs out of $(\Sigma Q_{ijk})^{2m}$ is $\frac{(2m)!}{2^m m!}$ so that the leading term in μ_{2m} is

$$\frac{(2m)!}{2^m m!}(\tfrac{3}{32})^m n^{3m},$$

whilst the leading term in μ_2 is $\frac{3}{32}n^3$ and so

$$\frac{\mu_{2m}}{\mu_2^m} \to \frac{(2m)!}{2^m m!}.$$

The distribution therefore tends to normality.

REFERENCES

KENDALL, M. G. & BABINGTON SMITH, B. (1940). *Biometrika*, **31**, 324.

KENDALL, M. G. (1943). *The Advanced theory of Statistics*, **1**. London: Charles Griffin and Co.

Notes on the calculation of autocorrelations of linear autoregressive schemes

BY M. H. QUENOUILLE

1. Bartlett (1946) has recently shown how, for a series of observations, we can test whether the observations can be adequately represented by a linear autoregressive scheme

$$u_{n+j}+a_1u_{n+j-1}+\ldots+a_ju_n = \epsilon_{n+j}, \tag{1}$$

where the a_i are known or fitted values, and ϵ_{n+j} is an error component independent of u_{n+j-1}. Bartlett's test is based on the formula

$$\operatorname{cov}(r_s, r_{s+t}) \sim \frac{1}{n-s}\sum_{i=-\infty}^{\infty}\rho_i\rho_{i+t},$$

where r_s is the estimate of the true autocorrelation ρ_s between u_i and u_{i+s}.

The purpose of the note is to demonstrate how, using generating functions, ρ_i and $\sum_{i=-\infty}^{\infty}\rho_i\rho_{i+t}$ can be calculated with the minimum of computation.

2. The method of generating functions seems to have been used by Wold (1938), who applied them to finding the variances and covariances of linear forms of finite extent in variables such as ϵ_{n+j}. We shall, however, be concerned with linear forms of infinite extent.

It can easily be shown that the solution of (1) can be written

$$u_n = \epsilon_n + b_1\epsilon_{n-1}+b_2\epsilon_{n-2}+\ldots, \tag{2}$$

where

$$(1+a_1t+\ldots+a_jt^j)^{-1} = 1+b_1t+b_2t^2+\ldots. \tag{3}$$

For example, if

$$u_{n+2}+au_{n+1}+bu_n = \epsilon_{n+2},$$

we have

$$(1+at+bt^2)^{-1} = (1-2x\cos\theta+x^2)^{-1} = 1+\frac{\sin 2\theta}{\sin\theta}x+\frac{\sin 3\theta}{\sin\theta}x^2+\ldots,$$

where

$$\cos\theta = -\tfrac{1}{2}a/\sqrt{b},\ x = t\sqrt{b},$$

and hence

$$b_l = b^{\frac{1}{2}l}\frac{\sin l\theta}{\sin\theta} = \frac{2b^{\frac{1}{2}l}}{\sqrt{(4b-a^2)}}\sin l\theta.$$

3. Using this generating function we have

$$\sigma^2 \sum_{i=-\infty}^{\infty} \rho_i t^i = \operatorname*{Lt}_{c \to 1} \left[\{1 + a_1 t + \ldots + a_j t^j\} \left\{1 + a_1 \frac{c}{t} + \ldots + a_j \left(\frac{c}{t}\right)^j\right\} \right]^{-1}. \tag{4}$$

Now the expansion of (4) can be achieved by splitting into partial fractions and, in general, we can let $c \to 1$ before this operation is performed. Thus

$$\frac{t^j}{(1+a_1 t + \ldots + a_j t^j)(t^j + a_1 t^{j-1} + \ldots + a_j)} = \frac{A_0 + A_1 t + \ldots + A_{j-1} t^{j-1}}{1 + a_1 t + \ldots + a_j t^j} + \frac{B_j + B_{j-1} t + \ldots + B_1 t^{j-1}}{t^j + a_1 t^{j-1} + \ldots + a_j}, \tag{5}$$

and using $\rho_i = -\rho_{-i}$, we can see that

$$A_i = -B_j/a_j \quad (i = 0)$$
$$= B_i - a_i B_j/a_j \quad (i = 1, \ldots j-1).$$

Thus the autocorrelations will be generated by

$$1 + \frac{t}{A_0} \frac{B_1 + B_2 t + \ldots + B_j t^{j-1}}{1 + a_1 t + \ldots + a_j t^j} + \frac{1}{A_0 t} \frac{B_1 + B_2 t^{-1} + \ldots + B_j t^{-j+1}}{1 + a_1 t^{-1} + \ldots + a_j t^{-j}}, \tag{6}$$

where the first term is expanded in powers of t and the second term is expanded in powers of t^{-1}.

4. The expression (6) can now be squared to give a generating function for $\sum_{i=-\infty}^{\infty} \rho_i \rho_{i+t}$. It will be necessary to split

$$\frac{t(B_1 + B_2 t + \ldots + B_j t^{j-1})(B_1 t^{j-1} + B_2 t^{j-2} + \ldots + B_j)}{(1 + a_1 t + \ldots + a_j t^j)(t^j + a_1 t^{j-1} + \ldots + a_j)}$$

into partial fractions, but the labour will be reduced since the matrix of the coefficients of the equations in B_i will be unaltered.

5. To illustrate the method, we can consider Kendall's series 1, which was used by Bartlett in his example.

The autoregressive scheme for this series is

$$u_{n+2} - 1{\cdot}1 u_{n+1} + 0{\cdot}5 u_n = \epsilon_{n+2},$$

so that $$\sigma^2 \sum_{i=-\infty}^{\infty} \rho_i t^i = \frac{t^2}{(1 - 1{\cdot}1t + 0{\cdot}5t^2)(t^2 - 1{\cdot}1t + 0{\cdot}5)} = \frac{2B_2 + (B_1 + 2{\cdot}2B_2)t}{1 - 1{\cdot}1t + 0{\cdot}5t^2} + \frac{B_2 + B_1 t}{t^2 - 1{\cdot}1t + 0{\cdot}5},$$

where $$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} -3{\cdot}7692 & 2{\cdot}1154 \\ -2{\cdot}1154 & -1{\cdot}4423 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2{\cdot}1154 \\ -1{\cdot}4423 \end{bmatrix}.$$

Thus $$\sigma^2 = 2{\cdot}8846,$$

and $$\sum_{i=-\infty}^{\infty} \rho_i t^i = 1 + t \frac{0{\cdot}7333 - 0{\cdot}5t}{1 - 1{\cdot}1t + 0{\cdot}5t^2} + \frac{1}{t} \frac{0{\cdot}7333 - 0{\cdot}5t^{-1}}{1 - 1{\cdot}1t^{-1} + 0{\cdot}5t^{-2}}, \tag{7}$$

so that $$\rho_i - 1{\cdot}1\rho_{i-1} + 0{\cdot}5\rho_{i-2} = 0 \quad (i > 0). \tag{8}$$

If we now consider the square of the expression (7) we have a product term

$$\frac{2t(0{\cdot}7333 - 0{\cdot}5t)(0{\cdot}7333t - 0{\cdot}5)}{(1 - 1{\cdot}1t + 0{\cdot}5t^2)(t^2 - 1{\cdot}1t + 0{\cdot}5)} = \frac{-2B_2 + (B_1 + 2{\cdot}2B_2)t}{1 - 1{\cdot}1t + 0{\cdot}5t^2} + \frac{B_2 + B_1 t}{t^2 - 1{\cdot}1t + 0{\cdot}5},$$

where $$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} -3{\cdot}7692 & 2{\cdot}1154 \\ -2{\cdot}1154 & -1{\cdot}4423 \end{bmatrix} \begin{bmatrix} -0{\cdot}7333 \\ 1{\cdot}5754 \end{bmatrix} = \begin{bmatrix} 0{\cdot}5686 \\ -0{\cdot}7210 \end{bmatrix},$$

and, if we write
$$P_t = \sum_{i=-\infty}^{\infty} \rho_i \rho_{i+t} \Big/ \sum_{i=-\infty}^{\infty} \rho_i^2,$$

$$\sum_{i=-\infty}^{\infty} \rho_i^2 \sum_{i=-\infty}^{\infty} P_i t^i = 1 + 2t \frac{0\cdot7333 - 0\cdot5t}{1 - 1\cdot1t + 0\cdot5t^2} + t^2 \left(\frac{0\cdot7333 - 0\cdot5t}{1 - 1\cdot1t + 0\cdot5t^2}\right)^2$$
$$+ 1\cdot4420 + t\frac{0\cdot5686 - 0\cdot7210t}{1 - 1\cdot1t + 0\cdot5t^2} + \text{terms in } t^{-1}$$
$$= 2\cdot4420 + t\frac{2\cdot0352 - 1\cdot7210}{1 - 1\cdot1t + 0\cdot5t^2} + t^2\frac{0\cdot5377 - 0\cdot7333t + 0\cdot25t^2}{(1 - 1\cdot1t + 0\cdot5t^2)}$$
$$+ \text{terms in } t^{-1}. \tag{9}$$

From this we have $\sum_{i=-\infty}^{\infty} \rho_i^2 = 2\cdot4420$ and the 'correlations' P_i of the correlations are 0·8334, 0·4321, 0·0006, ... Successive terms may be calculated using the relation
$$P_i - 2\cdot21 P_{i-1} + 2\cdot2 P_{i-2} - 1\cdot1 P_{i-3} + 0\cdot25 P_{i-4} = 0 \quad (i > 0). \tag{10}$$

The calculation of $\sum_{i=-\infty}^{\infty} P_i^2$, suggested by Bartlett, can also be made by this method, but it is more arduous, and the first few terms will give a good approximation.

7. The same method can be used to calculate the appropriate number of degrees of freedom for testing the correlation between two linear autoregressive schemes.

In general, if $E(u_i u_j) = \rho_{ij}\sigma^2$, $E(v_i v_j) = \rho'_{ij}\sigma'^2$ and $r = \dfrac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\left(\sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} v_i^2\right)}}$,

then
$$\operatorname{var} r \sim \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{ij}\rho'_{ij}}{\sum_{i=1}^{n} \rho_{ii} \sum_{i=1}^{n} \rho'_{ii}}.$$

For linear autoregressive schemes, $\rho_{ij} = \rho_{i-j}$, $\rho'_{ij} = \rho'_{i-j}$, and thus
$$\operatorname{var} r \sim \frac{n + (n-1)\rho_1\rho'_1 + \dots + \rho_{n-1}\rho'_{n-1}}{n^2}$$
$$\sim \sum_{i=-\infty}^{\infty} \rho_i \rho'_i / n.$$

Thus, provided n is large, r can be tested with $n \Big/ \sum_{i=-\infty}^{\infty} \rho_i\rho'_i$ degrees of freedom, and the calculation of $\sum_{i=-\infty}^{\infty} \rho_i\rho'_i$ can be made by the above method.

8. Finally, it is worth noting that, for autoregressive schemes involving m observables, it is possible to extend this method by the use of m parameters to calculate the correlations within and between the observables, provided that adequate estimates of the coefficients of the equations are available. In practice, however, the procedure will often be reversed, and estimates of the coefficients of the autoregressive schemes will be obtained by equating the theoretical and observed correlations.

REFERENCES

Bartlett, M. S. (1946). *J.R. statist. Soc. Suppl.* 8, 27.

Wold, H. (1938). *A Study in the Analysis of Stationary Time-series.* Uppsala.

Approximate formulae for the percentage points of the incomplete beta function and of the χ^2 distribution

BY D. HALTON THOMSON

Valuable 'Tables of percentage points of the Incomplete Beta Function' have been published in *Biometrika* (Thompson, 1941*a*) giving numerical values of percentage points at various probability levels between $P = 0{\cdot}995$ and $P = 0{\cdot}005$ for degrees of freedom $\nu_1 = 2q$ and $\nu_2 = 2p$ ranging up to 120, and with an accuracy of five significant figures. In the same volume, a 'Table of the percentage points of the χ^2 distribution' was also published (Thompson, 1941*b*) for values at the same probability levels and degrees of freedom ranging up to $\nu = 100$, and with an accuracy of six significant figures, thus supplementing the table of that function originally due to R. A. Fisher (Fisher & Yates, 1938).

Cases arise in practice where the tails of the frequency distribution of a large population are of special interest, thus involving (in the case of the beta function) values of $2p$ larger than 120, with a small $2q$, or vice versa. Harmonic interpolation between 120 and infinity, however, leads to substantial errors, as is found when the values of the percentage points x are expressed in terms of their tail values (x or $1-x<0{\cdot}5$). This Note shows that close approximations to such extreme values may be determined by using the χ^2 table as an auxiliary table to extend the Beta Function Tables in conjunction with certain simple alternative formulae. Comparisons within the range of the published Beta Function Tables are made indicating the degree of accuracy within that range. The accuracy of these formulae beyond that range increases rapidly with increasing $2p$ and decreasing $2q$ (and vice versa), so that they can be applied with confidence under such conditions.

The 'normalized' form of the Incomplete Beta Function, in the usual notation, is

$$P = I_x(p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)}\int_0^x x^{p-1}(1-x)^{q-1}\,dx, \tag{1}$$

in which, for a given P, $1-x(q,p)$ in the tables denotes the upper percentage point and $x(p,q)$ the lower percentage point.

It is known that, when p is large and q is small compared with p, this form tends towards the Incomplete Gamma Function

$$P = \frac{p^q}{\Gamma(q)}\int_0^t e^{-pt}t^{q-1}\,dt,$$

where $x(p,q) = e^{-t}$. This in turn may be transformed to the χ^2 distribution by putting $pt = [\chi^2_\nu(P)]/2$. For a given large p and small q, therefore, the percentage point in terms of χ^2 is given approximately by

$$x(2p,2q) \cong \exp\left[-\frac{\chi^2_{2q}(P)}{2p}\right], \tag{2}$$

where $2q = \nu$ in the χ^2 table. This expression gives the exact value of x, when $2q = 2$, but for larger $2q$ the error, which is consistently negative, increases rapidly with increasing $2q$ unless $2p$ is very large—much larger than 120. It is, therefore, of limited practical use. The following modifications were in consequence evolved.

APPROXIMATION A

Consider the constant of integration in the original form (1) which, when expanded, is

$$\frac{(p+q-1)(p+q-2)\ldots(p+1)p}{\Gamma(q)}.$$

Let the terms $q-1, q-2, \ldots, 1, 0$ be averaged; the constant as a first approximation then becomes

$$\frac{\{p+\frac{1}{2}(q-1)\}^q}{\Gamma(q)}.$$

The numerator suggests that a more accurate approximation for x would be obtained by substituting $p+\frac{1}{2}(q-1)$ in place of p in (2), thus leading to

$$x(2p,2q) \cong \exp\left[-\frac{\chi^2_{2q}(P)}{2p+q-1}\right]. \tag{A}$$

A comparison of the approximate values of x obtained from (A) with the exact values in the Beta Function Tables, for all probability levels between $P = 0{\cdot}995$ and $P = 0{\cdot}005$, shows that:

(*a*) The error is consistently positive, but much smaller than the negative error in (2); in other words the latter is slightly over-corrected.

(*b*) For a given p/q and varying P, the error is nearly constant; it is smallest at $P = 0{\cdot}995$ and increases gradually in the direction of $P = 0{\cdot}005$.

(*c*) For a given P, the error decreases rapidly with increasing $2p$ and/or decreasing $2q$.

(*d*) Provided that p/q is larger than 4, the value of x is within 0·5 % of the exact tail value; if p/q is larger than 10, the error is within 0·1 % of that value.

Approximation B

The exponent in (A) may be written

$$\frac{\chi^2_{2q}(P)}{2p+q-1} = \frac{\chi^2_{2q}(P)}{2q}\,\frac{2q}{2p+q-1} = \frac{\chi^2_{2q}(P)}{2q}\left[\frac{2}{2\left(\dfrac{2p+2q-1}{2q}\right)-1}\right].$$

The factor in square brackets is equivalent to the first term in the known expansion of the form

$$\log\left(\frac{n-1}{n}\right) = -2\left\{\frac{1}{2n-1}+\frac{1}{3(2n-1)^3}+\dots\right\},$$

where $n = (2p+2q-1)/2q$, which converges rapidly when n is large; i.e. when $2p$ is large compared with $2q$. The above exponent may therefore be written

$$-\frac{\chi^2_{2q}(P)}{2q}\log\left(\frac{2p-1}{2p+2q-1}\right),$$

which, when inserted in (A), leads to
$$x(2p,2q) \simeq \left(\frac{2p-1}{2p+2q-1}\right)^k, \qquad \text{(B)}$$

where $k = [\chi^2_{2q}(P)]/(2q)$.

A similar comparison with the Beta Function Tables, for the same range of probability levels, shows that:

(*a*) Approximation (B) gives generally more accurate values than (A), except when $2q$ is very small, in which case they are nearly identical.

(*b*) For a given p/q and varying P, the error is negligible in the vicinity of $P = 0{\cdot}25$; it increases negatively in the direction of $P = 0{\cdot}995$, and positively in the direction of $P = 0{\cdot}005$, the largest errors occurring at this level.

(*c*) For a given P, the error decreases rapidly with increasing $2p$ and/or decreasing $2q$.

(*d*) Provided that $(2p)^3/(2q)^2$ is larger than about 150, the values of x are within 0·5 % of the exact tail value; this implies that if $2p$ is larger than about 150, this degree of accuracy is attained even when p/q is as low as unity. If $(2p)^3/(2q)^2$ is larger than about 2000, the error is within 0·1 % of the exact value, which implies that if $2p$ is larger than about 120, this degree of accuracy is attained when p/q is as low as 4.

It will be observed that, when $2q = 2$, the formula does not revert exactly to (2), as is required by theory; but, unless $2p$ is also quite small, the error in the computed value of x is negligible.

The expansion of (B) leads to
$$x(2p,2q) \simeq e^{-v} - s(1+\tfrac{4}{3}s)\,v + sv^2,$$

where
$$v = \frac{\chi^2_{2q}(P)}{2p+2q-1} \quad\text{and}\quad s = \frac{q}{2p+2q-1},$$

thus demonstrating its analogies with Campbell's formula (C) below.

Adaptation of Campbell's formula

In a book concerned primarily with quality control, Simon (1941) quotes (without the proof) a formula, due to Campbell (1923), designed to determine the average number of defectives in a sample of n, starting from the known average number in an infinite sample. It is a particular application of the general problem now under consideration, namely, the approximate determination of the percentage points of

the Beta Function, starting from the corresponding known values for the χ^2 form of the Poisson exponential binomial summation. It is given in the following form:

$$\frac{a(c,n,P)-a(c,\infty,P)}{a(c,\infty,P)} = An^{-1}+\tfrac{1}{12}[14A^2+(3a+2)A+a]n^{-2}+\ldots, \tag{3}$$

where $a(c,n,P)$ = average number of defectives in which P is the probability of at least c defectives in a sample of n, $a(c,\infty,P)$ = average number of defectives in an infinite sample, $A = \frac{1}{2}(c-a-1)$, in which $a = a(c,\infty,P)$. (Simon quotes $a = (a,\infty,P)$, which is an evident misprint.)

If G denotes the value given by the formula, then

$$a(c,n,P) = a(c,\infty,P)(1+G),$$

so that $1+G$ is the factor by which the average number of defectives in an infinite sample must be multiplied to give that in a sample of n.

The change from Campbell's notation to the more familiar general notation is given by

$$a(c,n,P) = \{1-x(2p,2q)\}n, \quad a = a(c,\infty,P) = [\chi^2_{2q}(P)]/2,$$

where
$$n = p+q-1, \text{ and } c = q.$$

Let
$$u = a/n \quad \text{and} \quad r = (c-1)/(2n),$$

then
$$A = n(r-u/2).$$

By inserting this notation in (3) and rearranging the terms, the formula leads to

$$x(2p,2q) \cong 1-\left\{1+r\left(1+\tfrac{7}{6}r+\frac{1}{6n}\right)\right\}u+(1+\tfrac{11}{6}r)\frac{u^2}{2}-\tfrac{1}{6}u^3+\ldots,$$

which expression includes the first four terms in the expansion of e^{-u}.

Hence, for the determination of the percentage points $x(2p,2q)$, Campbell's formula may, in effect, be re-written as

$$x(2p,2q) \cong e^{-u}-r\left(1+\tfrac{7}{6}r+\frac{1}{6n}\right)u+\tfrac{11}{12}ru^2, \tag{C}$$

where
$$u = \frac{\chi^2_{2q}(P)}{2(p+q-1)} \quad \text{and} \quad r = \frac{q-1}{2(p+q-1)}.$$

For large $2p$ and small $2q$, the last two terms become negligible, in which case it reduces to

$$x(2p,2q) \cong e^{-u}-r(1+\tfrac{7}{6}r)u. \tag{C'}$$

Cochran's approximation

Cochran (1940), extending a method of Fisher's (1925), has introduced a useful approximation for the percentage points of the Incomplete Beta Function, *when both p and q are large*, his method being to determine a sufficiently accurate value of z, as used in Fisher's z-transformation.

If y is the normal deviate at probability level P, then for a given pair of arguments $2p$, $2q$, the following are first calculated, using Hartley's (1941) notation:

$$\lambda = \tfrac{1}{6}(y^2+3), \quad A = \frac{8pq}{2p+2q},$$

$$z = \frac{y}{\sqrt{(A-\lambda)}}+\frac{(\lambda-\frac{1}{6})(A-2p)}{pA}.$$

Hence, by Fisher's transformation,
$$x(2p,2q) \cong \frac{2p}{2p+2qe^{2z}}. \tag{D}$$

Comparison of formulae

Table 1 compares the various formulae for upper percentage points at an extreme probability level ($P = 0{\cdot}995$). Table 2 indicates their relative accuracy on a common basis, namely, as a percentage of the exact value of x or $1-x$, whichever is the smaller, so that the deviations from the exact values, when x or $1-x$ approach zero, are duly emphasized. For intermediate probability levels, the percentages lie between the tabulated extremes. It will be noted that in the case of approximation B the errors pass through zero near the mid-range of P; in the cases of A and C the errors are positive for all values of P.

The general conclusions from these tables and other comparisons are that, for a given probability level P:

(*a*) When $p/q >$ about 6, approximations A, B and C have about the same degree of accuracy, so that the simpler, A or B, have the advantage.

(*b*) In the range $6 > p/q > 4$, there is little to choose between B and C; but B is the simpler.

(*c*) When $p/q < 4$ and the distribution approaches symmetry, D gives the best results, provided that $2p$ and $2q$ are moderately large, say > 50. It may be, however, that B in this range will be sufficiently accurate for many purposes; if $p/q > 2$, the maximum error of x is about 2 units in the third decimal place.

Table 1. *Comparison of approximate formulae at a given probability level*

$P = 0{\cdot}995$

$2p$	$2q$	$x(2p, 2q)$				
		A	B	C (Campbell)	D (Cochran)	Exact
120	2	$0{\cdot}9^{4}16461$ *Nil*	$0{\cdot}9^{4}16459$ $-0{\cdot}0^{4}00002$	$0{\cdot}9^{4}16461$ *Nil*	0·999862 *−0·000054*	$0{\cdot}9^{4}16461$ —
	4	0·9982908 *+0·0000001*	0·9982906 *−0·0000001*	0·9982907 *Nil*	0·997926 *−0·000365*	0·9982907 —
	10	0·982764 *+0·000005*	0·982755 *−0·000004*	0·982760 *+0·000001*	0·982076 *−0·000683*	0·982759 —
	20	0·944002 *+0·000072*	0·943893 *−0·000037*	0·943941 *+0·000011*	0·943366 *−0·000564*	0·943930 —
	30	0·902230 *+0·000280*	0·901839 *−0·000111*	0·902000 *+0·000050*	0·901551 *−0·000399*	0·901950 —
	40	0·86160 *+0·00067*	0·86070 *−0·00023*	0·86106 *+0·00013*	0·86066 *−0·00027*	0·86093 —
	60	0·78782 *+0·00203*	0·78522 *−0·00057*	0·78621 *+0·00042*	0·78568 *−0·00011*	0·78579 —
	120	0·62598 *+0·00978*	0·61430 *−0·00190*	0·61855 *+0·00235*	0·61620 *Nil*	0·61620 —

N.B. The figures in italics are the differences between the approximate and exact values.

Table 2. *Relative accuracy of approximate formulae*

Error of $x(2p, 2q)$ expressed as a percentage of the smaller exact tail value (x or $1 - x < 0{\cdot}5$)

$2p$	$2q$		A			B			C (Campbell)			D (Cochran)		
		p/q \ P	0·995	0·500	0·005	0·995	0·500	0·005	0·995	0·500	0·005	0·995	0·500	0·005
			%	%	%	%	%	%	%	%	%	%	%	%
120	12	10	*	*	*	*	*	*	*	*	*	−2·8	−0·1	+0·3
	20	6	+0·1	+0·2	+0·2	−0·1	*	+0·1	*	*	+0·1	−1·0	*	+0·1
	40	3	+0·5	+0·6	+0·7	−0·2	*	+0·2	+0·1	+0·1	+0·3	−0·2	*	*
	60	2	+0·9	+1·1	+1·2	−0·3	*	+0·2	+0·2	+0·1	+0·5	−0·1	*	*
	120	1	+2·5	+2·7	+4·4	−0·5	*	+0·7	+0·6	+0·5	+1·7	*	*	*
30	3	10	*	*	+0·1	*	*	*	*	*	+0·1	−30·0	−1·4	−7·1
	5	6	+0·1	+0·1	+0·3	−0·1	−0·1	+0·1	*	*	+0·4	−11·1	−0·5	−1·8
	10	3	+0·4	+0·5	+0·8	−0·3	−0·1	+0·3	+0·1	+0·1	+1·3	−2·2	−0·1	−0·5
	15	2	+0·8	+1·0	+1·9	−0·5	−0·1	+0·6	+0·3	+0·1	+2·8	−0·7	*	−0·3
	30	1	+2·4	+2·7	+7·1	−1·0	−0·1	+1·8	+1·0	+0·5	+8·2	*	*	*

* Error smaller than ± 0·05 %.

Wilson-Hilferty approximation for χ^2-adjustment

This formula (Wilson & Hilferty, 1931) for the percentage points of the χ^2 distribution is

$$\chi^2_\nu(P) = \nu\left\{1-\frac{2}{9\nu}+y_P\sqrt{\frac{2}{9\nu}}\right\}^3,$$

where ν represents the degrees of freedom, and y_P the standardized normal deviate corresponding to probability level P. A table has been published in *Biometrika* (Merrington, 1941), comparing the approximations derived from this formula with the exact values, at various probability levels between $P = 0{\cdot}995$ and $P = 0{\cdot}005$. It shows the remarkable accuracy of the formula, the maximum errors varying from about $\pm 0{\cdot}04$, when $\nu = 30$, to about $\pm 0{\cdot}024$, when $\nu = 100$.

When these errors were plotted against the exact values on logarithmic paper, it was observed that for a given probability level, they varied inversely with $\sqrt{\nu}$ very closely. It follows that this square root relation may be used to adjust the Wilson-Hilferty formula, bringing the values computed therefrom still nearer to the exact values.

If the difference (at $\nu = 30$) between the Wilson-Hilferty value and the exact value, when multiplied by $\sqrt{(100/30)}$, is treated as a coefficient C (which may be positive or negative), the required adjustment for any value of ν is given by

$$\text{Adjustment} = C/\sqrt{\nu}.$$

For various probability levels P, the values of C are given in the following table:

P	C	P	C
0·995	+0·233	0·250	+0·039
0·990	+0·157	0·100	+0·056
0·975	+0·067	0·050	+0·035
0·950	+0·011	0·025	−0·015
0·900	−0·029	0·010	−0·120
0·750	−0·046	0·005	−0·227
0·500	−0·013	—	—

A test against the Merrington Table shows that this adjustment leads to values of χ^2, between $\nu = 30$ and $\nu = 100$ at all probability levels with an accuracy of $\pm 0{\cdot}001$, i.e. to four or five significant figures. Since the Wilson-Hilferty approximation assumes a normal distribution about $1-2/(9\nu)$, which tends to unity as ν increases to infinity, and since the adjustment tends to zero under those conditions, it follows that the latter may also be safely applied for an indefinitely large ν.

It should be added that an adjustment on similar principles is not applicable to the Fisher approximation for χ^2.

REFERENCES

Campbell, G. A. (1923). *Bell Syst. Tech. J.* January.

Cochran, W. G. (1940). Note on an approximate formula for significance levels of z. *Ann. Math. Statist.* **11**, 93.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd. 1st edition.

Fisher, R. A. & Yates, F. (1938), *Statistical Tables for Biological Agricultural and Medical Research.* Edinburgh: Oliver and Boyd.

Hartley, H. O. (1941). Tables of percentage points of the Incomplete Beta Function. Methods of interpolation. *Biometrika*, **32**, 166.

Merrington, M. (1941). Numerical approximations to the percentage points of the χ^2 distribution. *Biometrika*, **32**, 200.

Simon, L. E. (1941). *An Engineer's Manual of Statistical Methods*, p. 185. New York: John Wiley and Sons, Inc.

Thompson, C. M. (1941*a*, *b*). Tables of percentage points of the Incomplete Beta Function, *Biometrika*, **32**, 151.

Thompson, C. M. (1941). Table of percentage points of the χ^2 distribution. *Biometrika*, **32**, 187.

Wilson, E. B. & Hilferty, M. M. (1931). The distribution of chi-square. *Proc. Nat. Acad. Sci. Wash.* **17**, 684.

REVIEWS

A First Course in Mathematical Statistics. By C. E. WEATHERBURN. Cambridge University Press. Price 15*s*.

An outstanding feature of the present statistical time is the number of text-books which are being written, and each one from a slightly different point of view. It is this which makes statistical theory interesting to study, for there can be no rigid approach to a subject which is used and expounded by so many and diverse persons. Professor Weatherburn has taken a rather formal mathematical exposition of the subject, and mathematical students will find his book both interesting and profitable to read. Numerical examples are given for the reader to apply the appropriate mathematical technique. It is possible that these would have been of greater utility if they had contained the material in its crude state, and had not been streamlined so that the application of the technique is immediately obvious, but nevertheless many new examples are there.

I am not sure whether this book will be entirely useful to students of other subjects than mathematics. While the mathematical analysis is undoubtedly clear it is possible that many will not be able to follow it in detail, and the conclusions of the analysis are not emphasized strongly. We may contrast with this Fisher's *Statistical Methods for Research Workers*, where no analysis is given, but where the relevant formulae and their interpretation are stated unmistakeably and their applications to material in its crude state set out so that the student may calculate for himself.

Probability theory is the foundation stone on which the whole of statistical theory is built. It is disappointing therefore to find that it is given somewhat perfunctory treatment in one chapter and the part it plays in (say) statistical tests of significance is not brought out and emphasized. There is a tendency nowadays in applying statistical technique to regard the 5 % and 1 % levels of significance as sacrosanct and those coming fresh to the subject should learn that custom is the only reason for their choice.

In spite of the criticisms which I make, however, I would recommend this book to students who have obtained some idea of the aims and objectives of statistical theory, and who are desirous of learning the development of the mathematical technique as well as its application. Professor Weatherburn's mathematical analysis makes pleasant reading and may well throw new light on old methods for those who have learnt the rudiments of the theory.

F. N. DAVID

Advances in Genetics, Volume 1. New York, N.Y.: Academic Press. 1947.

This is the first number of a new periodical, probably an annual, summarizing recent work in various fields. Of the nine articles, ranging from 12 to 96 pages, with mean 42·6, S.D. 7·89, and a positively skew distribution, perhaps the most interesting to European geneticists will be that on the genetics of the ciliate Protozoa, *Paramecium* and *Euplotes*. Here Sonneborn describes work almost entirely done in America, with very surprising results. Thus *Paramecium aurelia* consists of at least seven endogamous varieties, each with two exogamous mating types, which might be called sexes were it not that in *P. bursaria* one of the varieties has no less than eight mating types.

Shrode and Lush's article of the genetics of cattle gives a very condensed account of the large amount of work which has been done on the inheritance of economically important characters such as milk yield and growth rate. For example cattle biometricians have used the important concept of 'heritability', meaning the fraction of the variance of a character due to additive genetic differences. Within a herd this rarely exceeds 30 %. More space is devoted to work on the genetics of colour and the like, which is of far less economic importance, and the review of progeny testing methods is disappointingly brief. However, the bibliographical references will be useful. Similarly, Atwood's article on forage crops, though most valuable as a guide to the literature, does not give a detailed account of any of the biometric work which has been done on grasses and clovers.

Only two of the papers give data which a biometrician could immediately utilize. These are Gordon's account of polymorphism in fish populations, and Spencer's of mutations in wild *Drosophila* species, which unfortunately does not include some valuable recent Italian and Russian work. Gordon's results call for the development of methods of estimating gene frequency similar to those used with human blood groups. Spencer is mainly concerned with results, but these are often given in sufficient detail to interest biometricians, though no attempt is made to summarize Wright's fundamental statistical theory.

The other articles will be less attractive to biometricians, though it is of interest to see how statistical methods are demanded by the mere fact that the genus *Crepis*, whose evolution is reviewed by Babcock, includes 196 species, most of which have been examined cytologically, and between which 130 of the 38,220 possible crosses have been made.

The volume will be indispensable to geneticists. Biometricians certainly cannot neglect it.

J. B. S. H.

Mathematical Methods of Statistics. By H. Cramér, Princeton University Press. 1946. $6.00.

This book was written by Prof. Cramér during the war and has been published first in Sweden and then by an offset process by the Princeton University Press in the U.S.A. It is a definitive exposition of the theory of mathematical statistics as it existed in 1940 (about) and it is worth while therefore to consider its contents in some detail. Prof. Cramér has divided his exposition into three parts; the first part is purely mathematical. The theory of sets and of such Lebesgue measure as is necessary for the understanding of the second part is developed first of all. Such a development will be useful for the student of mathematical statistics coming fresh to the theory of measure in that he receives guidance as to what are the elements essential for him to understand. Chapters 11 and 12 on matrices determinants and quadratic forms and miscellaneous complements do not fit into this general scheme but have obviously been included here as part of the mathematical equipment necessary for the student. Possibly Chapter 10 on Fourier Integrals would have fitted more naturally into Part II but this is a matter of taste.

Part II begins with a formal development of the theory of probability as given by the French and Russian schools of probability, and which Prof. Cramér has already given in his Cambridge tract 'Random Variables and Probability Distributions'. The treatment here seems simpler, however, than in his earlier tract and there is a more practical flavour to his exposition. This part while still purely mathematical begins to introduce distributions and ideas which are familiar to the statistician.

The title of the third part is 'Statistical Inference' and the main outline is that of small sample theory developed during the past twenty-five years. The illustrations are numerical as well as mathematical and an attempt is made to show the student the numerical applications of the processes through which his mathematical theory leads him. The treatment is not exhaustive but the student who has assimilated this part will have little difficulty in extending his knowledge by further reading.

As a textbook of mathematical statistics this book will remain unrivalled for many years to come. The mathematical exposition is clear, the development of ideas logical throughout, and the theorems are presented in a very general way. Any student of mathematics who wishes to get a picture of what statistical theory is about will be led inevitably to a study of this book. To those who wish to become statisticians it will be necessary to supplement the reading by a practical course in which the mathematical tools are tried out on numerical examples. This aspect of statistical work the book does not cover, but it is obvious that this would be the case from the title. It only remains to say to the student 'This is a good book, buy it'.

F. N. DAVID

CORRIGENDA

(*Biometrika*, **34**, 176–7)

In J. Wishart's paper on 'The cumulants of the z and of the logarithmic χ^2 and t distributions', the following correction should be made:

p. 176, 1st line of section 3: *read* '$\log|t|$' *for* '$\log t$', in two places.

p. 177, 1st line following equation (32): *read* '$\log|x|$' *for* '$\log x$'.

BIOMETRIKA. Vol. XXXIV, Parts III and IV

CONTENTS

A volume of *Biometrika* contains about 400 pages, with plates and tables, and it is hoped that in future this will be published annually in two half-yearly issues.

Papers for publication should either be sent to

PROFESSOR E. S. PEARSON, Department of Statistics, University College, London, W.C. 1,

or if more convenient may be submitted through a member of the Editorial Committee, viz.

It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted without leave of the Editors.

Contributors receive 25 copies of their papers free. Joint authors 15 copies each.

The subscription price, *payable in advance*, is Inland 45*s. net* per volume and Abroad 54*s. net* (including packing and postage). Owing to the scarcity of early volumes, the following rates must now be charged for complete sets. Vols. I—XXXIV, including XX^B: £126. 5*s.* in wrappers, not including postage. At present certain volumes are out of print, but steps are being taken to re-issue these as quickly as printing facilities permit. Recent volumes may still be obtained at the wrapper price; this is 64*s.* inland, including postage. Index to Vols. I to V, 2*s. net*. Index to Vols. I to XV, 5*s. net*. Cheques must be made payable to *Biometrika*, crossed "*a/c Biometrika Trust*" and sent to The Secretary, *Biometrika Office*, Department of Statistics, University College, London, W.C. 1, to whom all orders for series, single copies and offprints should be addressed. All foreign cheques must be drawn in sterling and on a Bank having a London Agency.

First printed in Great Britain at the University Press, Cambridge

Reprinted by offset-litho by Percy Lund Humphries & Co., Ltd.